



Delft University of Technology

Human factors of monitoring driving automation

Eyes and Scenes

Cabral, Christopher

DOI

[10.4233/uuid:9e5bc7a2-fc0e-47ef-aee3-f0fce4809e5c](https://doi.org/10.4233/uuid:9e5bc7a2-fc0e-47ef-aee3-f0fce4809e5c)

Publication date

2019

Document Version

Final published version

Citation (APA)

Cabral, C. (2019). *Human factors of monitoring driving automation: Eyes and Scenes*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:9e5bc7a2-fc0e-47ef-aee3-f0fce4809e5c>

Important note

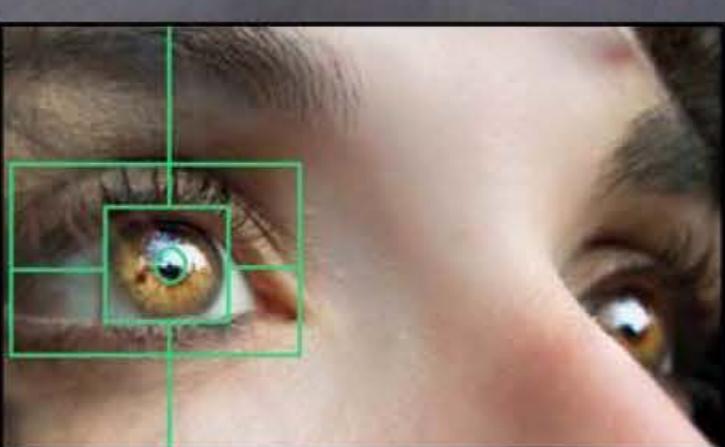
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



```
2019-22-05 15:22:44.215 exec driver_monitor_system
2019-22-05 15:22:44.325 init -lvl:2 -context:y
2019-22-05 15:22:45.192 param -fixation_dur: 0.235
2019-22-05 15:22:45.198 param -fixation_xy: -137 +441
2019-22-05 15:22:46.235 param -avg_blink_dur: 0.329
2019-22-05 15:22:46.385 param -eye_eccentricity: 5.137
2019-22-05 15:22:46.572 param -veh_velocity: 37.05
2019-22-05 15:22:46.836 param -road_surface: 0.972
2019-22-05 15:22:46.986 param -road_edge: 0.085
2019-22-05 15:22:47.078 param -road_grade: 0.357
2019-22-05 15:22:47.172 param -road_curve: 0.238
2019-22-05 15:22:47.312 param -traffic_density: light
2019-22-05 15:22:47.761 param -hands_on_wheel: 1.0
2019-22-05 15:22:47.828 param -posture: relaxed
2019-22-05 15:22:47.835 param -vigilance_thresh: 25 mins
```

Human Factors of Monitoring Driving Automation:

Eyes and Scenes

Christopher Donald Douglas Cabrall

Human Factors of Monitoring Driving Automation: Eyes and Scenes

Christopher D. D. Cabrall

Human Factors of Monitoring Driving Automation:
Eyes and Scenes

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof. dr. ir. T.H.J.J. van der Hagen,
Chair of the Board for Doctorates
to be defended publicly on
Monday 08 July 2019 at 10:00 o'clock

by

Christopher Donald Douglas CABRALL
Master of Science in Human Factors and Ergonomics, San Jose State University,
California, USA
born in Berkeley, California, USA

This dissertation has been approved by the promotores:

Prof. dr. F.C.T. Van der Helm
Dr. ir. J.C.F. De Winter
Dr. ir. R. Happee

Composition of the doctoral committee:

Rector Magnificus	Delft University of Technology, chairperson
Prof. dr. F.C.T. Van der Helm	Delft University of Technology, promotor
Dr. ir. J.C.F. De Winter	Delft University of Technology, promotor
Dr. ir. R. Happee	Delft University of Technology, promotor

Independent members:

Prof. dr. J. Flach	Wright State University
Prof. dr. F. Flemisch	RWTH Aachen University / Fraunhofer FKIE
Prof. dr. M.H. Martens	University of Twente
Prof. dr. ir. D. Abbink	Delft University of Technology
Prof. dr. ir. J. Hellendoorn	Delft University of Technology, reserve member

This research has been financially supported by the Marie Curie Innovative Training Network (ITN) through the project Human Factors of Automated Driving (HFAuto), grant PITN-GA-2013-605817.



Keywords: automated driving, eye tracking, traffic safety, human computer interaction, human supervisory control

Printed by: Gildeprint

Cover by: Christopher Cabrall

Copyright © 2019 by Christopher CABRALL

PhD series

ISBN 978.94.6323.714.7

An electronic copy of this dissertation is available at <http://repository.tudelft.nl/>

Table of Contents

Summary.....	viii
Samenvatting.....	xiii
PART 1: Introduction.....	1
1. Background.....	2
2. Driver Monitoring Systems.....	5
3. Theoretical Framework.....	7
4. Thesis Aims.....	11
5. Thesis Structure.....	12
References.....	14
PART 2: Driver Vigilance Review.....	17
Chap. 2.1) From Mackworth's Clock to the Open Road: A literature review on driver vigilance task operationalization (2016) <i>Transportation Research Part F</i>	19
Abstract	20
1. Introduction.....	21
2. Methods	26
3. Results	33
4. Discussion	44
5. Summary and Limitations	47
6. Directions for Future Research.....	48
References.....	50
Appendix A. Supplementary Material.....	55
Chap. 2.2) How to Keep Drivers Engaged while Supervising Driving Automation? A literature survey and categorization of six solution areas (in press) <i>Theoretical Issues in Ergonomics Science</i>	57
Abstract	58
1. Background.....	59
2. Introduction of Solution Grouping Framework	62
3. Methods of Literature Survey.....	70
4. Results of Rater Categorizations.....	71
5. Discussion	76
6. Conclusions.....	80
References.....	83
Appendix A. Literature Survey List.....	90
Appendix B. Category Ratings.....	92

PART 3: Driving Scenes and Driver Eyes.....	93
Chap. 3.1) Validity and Reliability of Naturalistic Driving Scene Categorization Judgments from Crowdsourcing (2018) <i>Accident Analysis & Prevention</i>	95
Abstract	96
1. Introduction.....	97
2. Methods	98
3. Results	104
4. Discussion, Conclusions and Recommendations	108
References.....	112
Appendix A. Coding Instructions/Training Material	113
Appendix B. Developed Driving Research Tools	131
Chap. 3.2) Estimating Driver Readiness from Situated Eye Movements: Prediction of workload and attention requirements from quantification of driving scene components (under review)	
T.B.D	133
Abstract	134
1. Introduction.....	135
2. Methods	142
3. Results	145
4. Discussion.....	151
5. Conclusions.....	155
References.....	157
Appendix A. Developed Driving Research Tools	161
Chap. 3.3) On-road Driver vs. Passenger Eye Eccentricity in a Conventional Car for In- vs. Out-of- the-loop “Drivenger” Monitoring in Automated Vehicles (under review) T.B.D	163
Abstract	164
1. Introduction.....	165
2. Methods	172
3. Results	175
4. Discussion	180
5. Application.....	182
Key points	183
References.....	184
PART 4: Adaptive Driving Automation	189
Chap. 4.1) Adaptive automation: Automatically (dis)engaging automation during visually distracted driving (2018) <i>PeerJ Computer Science</i>	191
Abstract	192
1. Introduction.....	193
2. Methods	195

3. Results	202
4. Discussion	208
5. Conclusions and Recommendations.....	213
References.....	214
Chap. 4.2) Redesigning Today's Driving Automation Towards Adaptive Backup Control with Situated and Implicit Interfaces (under review) <i>T.B.D.</i>.....	217
Abstract	218
1. Introduction.....	219
2. Methods	222
3. Results	231
4. Discussion	239
Key points	244
References.....	245
Appendix A. Developed Driving Research Tools	249
PART 5: Discussion	253
1. Thesis Imperatives and Impact Overview	254
2. Summary and Connection of Thesis Study Conclusions	255
3. Recently Convergent Research.....	260
4. Thesis Research and Development Implications	263
5. Future Research and Recommendations	266
6. Conclusion	269
References.....	270
Acknowledgments.....	272
List of Publications	275
Theses.....	275
Journal publications.....	275
Conference proceedings	276
Curriculum Vitae	282
Propositions Belonging to the PhD Thesis	283

Summary

Problem Statement

The World Health Organization recognizes road crashes as a public health epidemic with daily fatalities averaging over 100 in the USA and more than 3,000 worldwide. In the search for underlying causes to address, human error (particularly that of inadequate attention) is commonly identified as a principal culprit. Consequently, today's automotive industry and its safety advocates are very keen on advancing an automated/autonomous vehicle (AV) agenda to transform the domain. However, a human factors complication arises by releasing AV technology onto publics roads in an evolving manner: the continuous driving task changes to a monitor and fallback for driving automation. Generally, human operators are expected to face challenges for sustaining attention in particular for the transitional stages of the SAE levels of driving automation: whether they are end-consumers serving as full-time supervisors (SAE level 2) or on-call backups (SAE level 3), or hired test drivers over-seeing the development of autonomous vehicles (SAE level 4+).

Thesis Aim and Approach

Within a larger Human Factors of Automated Driving project (HFAuto, PITN-GA-2013-605817), the objective of this thesis was: *'to develop a system that is able to monitor the driver's vigilance'*. With an Oxford English Dictionary definition of *'the action or state of keeping careful watch for possible danger or difficulties'*, vigilance is thus entailed in all kinds of driving. However, because driving does not actually require full-time and undivided conscious attention (despite contrary casual assumptions), practical problems immediately appear when attempting to operationalize *'careful'*, *'danger'*, and/or *'difficulty'* and especially for driver monitor systems (DMS) where unnecessary alerts degrade end-user trust, acceptance, and adherence to the system (*'the cry-wolf effect'*). More knowledge of specific driving attentional requirements (i.e., how much under what circumstances) is expected to produce better assessments of the readiness of drivers across levels of driving automation.

The selected approach to meet the given thesis objective was to investigate vigilance from a cognitive systems engineering approach (ecological perspective). Instead of restricting the concept of vigilance to be some kind of internal state/property of a driver, this thesis treated vigilance as a state/property of a system (i.e., the relationship between a driver and a driving scene/situation). To differentiate from the traditional status quo, this thesis purposefully prepends the qualifier *'situated'* to describe cognition, vigilance, and/or DMS, etc. that directly takes into account present circumstances (the driving scene) in conjunction with conventional driver-centric measures/constructs.

Methods

Recently, video recording and processing technology have undergone exponential gains in capability with reduced form factors and costs. Thus, camera-based physiological and environmental measures (esp., eye and scene tracking/segmentation) should be increasingly useful research application areas to support a cognitive systems engineering approach of situated vigilance monitoring for driving. Increasing levels of AV control diminish hands-on and feet-on activity as sources of information about a driver's present behaviors, so videos (and eye tracking) remain as viable sources for driver assessment. Research/application questions progress from the

like of '*What is the driver's attention/vigilance level?*' to concerns more akin to '*Is the driver engaged/vigilant enough for the present demands?*' by simultaneously considering filmed aspects of the driving scene (and relating task contents and demands). Upon detecting imbalances, a situated driver monitor system functions to restore nominal balance between driver and scene demands via various kinds of DMS involvement, whether of information (notices, warnings, alarms, etc.) and/or actions (deceleration, transition of control, etc.).

Because accurate representations of real-life work domains and ecological constraints are essential to cognitive systems engineering approaches, this thesis recorded and related different eye measurements of both nominal and aberrant visual control, under a variety of high/low demand driving conditions from both in the lab and out on the road. Thus, the present thesis included a range of low, medium, and high fidelity methods to investigate situated applications of driver eye measurement towards issues of vigilance assessment. Across the thesis, theoretical and empirical research was used in the form of literature survey/review, non-intrusive eye-tracking measures, dash-cam driving scene film recordings, crowdsourced driving scene content categorizations, on-road measurements and a driving simulator.

Results and Connections

This thesis consists of five parts; the first part introduces relevant background theory and the framework underlying the thesis and the last part discusses major conclusions. Parts 2-4 focus on reviews for driver vigilance (Part 2), experiments to relate driving scenes and driver eyes (Part 3), and the integration of eye-based DMS with adaptive driving automation in a driving simulator (Part 4).

Chapter 2.1 aimed to characterize vigilance tasks applied in driving research, in terms of instructions/conditions, signal types/rates, and work component features in comparison to the classic vigilance literature. The review supported the importance of vigilance tasking details (i.e., 18 are provided in Table 2.1.1) that are lacking for predicting/managing conventional driving vigilance decrement situations: specific consensus definitions of conventional driving signal(s), noise, and required responses. However, for supervising automated driving, properties in common with classic vigilance decrement theory were discussed as increasing the likelihood of problems: temporal and spatial uncertainty of intermittent/rare signals requiring time critical response, within prolonged task durations and increased monotony. Conclusions from Chapter 2.1 thus recommended caution and suggested (re)design opportunities for deploying automated driving.

Chapter 2.2 proposed six solution area themes to problems of vigilance decrements in human supervision over automation. Generally, the first three themes described avoidance strategies either in a hard sense or different versions of a soft stance: objective or subjective supervisory control task reductions. The latter three themes were based from general learning theory paradigms in a chronological order: behaviorism, cognitivism, and ecological constructivism. Specifically, the solution areas were enumerated, labeled, and exemplified as follows. Solution Area (1): Avoid the role of sustained human supervision of automation (i.e., suspend/repeal/skip levels of automation requiring human oversight and backup). Solution Area (2): Reduce the supervising role along an objective dimension (i.e., change the amount of time or envelope of automated operations). Solution Area (3): Reduce the supervising role along a subjective dimension (i.e., share responsibilities and/or alter the end user experience and impressions). Solution Area (4): Support

the supervising role from the behaviorism paradigm (i.e., condition the desired target behaviors through training and/or selection). Solution Area (5): Support the supervising role from the dyadic cognitivism paradigm (i.e., inform designs to support cognitive processes and mental models). Solution Area (6): Support the supervising role from the triadic ecological paradigm (i.e., inform designs to leverage external environment contexts and/or task considerations).

Results from Chapter 2.2 showed that independent raters were able to reliably apply the themes to categorize recommendations from influential human-automation interaction research. The most common solution areas to the problem of keeping attention while supervising automation included those focused on internal cognitive states, followed by those with a broader situational (task/ecological) perspective.

Taken together, the studies of Part 2 emphasize the importance of cognitive and situational themed approaches for managing vigilance issues in general, but a lack of available practical details (i.e., what driving scene features and driver eye measurements) with which one might proceed to build a situated DMS. Thus, applied driver eye and driving scene measurement studies were conducted in Part 3.

Chapter 3.1 produced a broad yet efficient driving scene content categorization scheme for feature presence/absence (Appendix 3.1.B) e.g., type and locations of other road users, vehicular behavior such as lane changes and turns, and infrastructural details like road-markings, signage, and road curvature, etc. Chapter 3.1 confirmed relatively high levels of accuracy and reliability in crowdsourced annotations using that scheme. Because external crowdworkers completed the scene categorizations about ten times faster than conventional internal confederate researchers without degradation in the quality of that work, crowdsourcing is considered to offer compelling potential to situational driving safety research. Overall, measurement of driving scene aspects was nailed down in a concrete and viable manner which suggest that contextualized driving information is not to nebulous/arduous to collect and capture.

Chapter 3.2 determined specific driving scene features (i.e., road curvature and traffic) to be of importance to perceived driving effort ratings and associated behavioral, rather than cognitive, eye movements (i.e., saccade amplitude). The high volume of annotated scene segments in Chapter 3.1 (~12,862 scenes from around 50 different driving videos) enabled a selection of stimulus material that contained a sufficient degree of resolution to perform predictive regression analyses in Chapter 3.2 (i.e., continuous scaled independent variables to match continuous scaled dependent variable constructs). For example, one of the resultant equations represents the amount of perceived effort to expect in the presence of specific amounts of driving scene contents, while another, the consequential amount of saccade amplitude. Notably, the lower level eye movement measurements showed stronger (more reliable) relations with perceived effort and visible scene contents (lateral/longitudinal conflicts) than the higher level representation (and eye measurement) aspects of information uptake (fixation duration) and increased cognitive processing (pupil size).

Chapter 3.3 measured both on-road eye movements and driving scene aspects. '*Out-of-the-loop*' eyes generally exhibited greater off-center movement distances across entire trips. However, the off-center distances of '*in-the-loop*' eyes were observed to periodically rise and fall with

respectively low and high driving scene demands (as operationalized by steering angle, traffic count, and speed).

Taken together, the studies of Part 3 emphasize the viability of measuring relations between driver eyes and driving scenes at a behavioral level. An applicable situated DMS conclusion was that specific measureable (visible) scene demand features of road curvature and traffic count could reliably be represented in low-level pre-cognitive eye movement measurements. Next, the studies of Part 4 executed simulator proof-of-concept design validations of various integrations of real-time vigilance DMS and driving automation.

Chapter 4.1 implemented a driving simulator proof-of-concept real-time DMS and driving automation integration (i.e., where the automation backs up a driver that looks away too long) that showed safety and acceptance improvements over an emulated concept of present-day on-market functional allocations of automated driving (i.e., where the automation de-activates itself upon detecting distraction).

Chapter 4.2 extended the successful proof-of-concept from Chapter 4.1. Inattention problems with supervising driving automation were evidenced (but also reduced from a condition requiring one hand be kept on the wheel). Situated and implicit DMS integration designs of adaptive-backup control showed user interaction and performance improvements.

Taken together, the studies of Part 4 emphasize problems with presently released driving automation designs where humans supervise without continuous physical involvement requirements. Most importantly, the Part 4 studies confirm viability of real-time eye-based DMS integration with driving automation towards practical user experience and safety advantages not only when deployed in an adaptive-backup directionality for transition of control, but also as from a situated version of DMS specifically.

Conclusions, Recommendations, and Impact

It can be concluded from this thesis, that to develop DMS of driving vigilance, eye measurements (especially of movement distances) and scene contents (especially road curvatures and collision hazards) are important and relatable factors. Furthermore, it is concluded that these factors are obtainable in viable ways for future research and development application efforts. Specifically, the present thesis studies suggest means for DMS to be targeted to protect and maintain the lower foundational level or inner-most loop of driving attention at a behavioral level (rather than interactive implicit cognitive layers and representational experiences that can be added on top).

To achieve automatic DMS contributing to transportation safety we need to include human-like intelligence in DMS assessments of human beings across levels of driving automation. Humans are an adaptive and social species that take/expect situated information and judgments as a given (esp. when they are being criticized as being negligent). While retaining meaningful specificity that avoids misses, perceived false alarms from end-users should be reduced by DMS use of lower-level behavioral (visuomotor) assessments of eyes and scene features taken together in relation to each other. Practical recommendations for future research fall under two general categories: (1) greater fidelity/complexity in driving simulations (e.g., more traffic, intersections, and real-life secondary

tasks should provide greater generalizability of naturalistic driver adaption to driving scene demands) and (2) greater instrumentation technology in on-road vehicles (e.g., better knowledge of the driving scene contents and eye movement behaviors with improved measurement capabilities). Additionally, driving video recordings are recommended as a growing research resource that offers a hybrid of enhanced stimulus/behavioral fidelity towards on-road applications that also allows for laboratory levels of repeatability and control.

A situated approach is expected to better avoid cognitive ambiguity/dilemmas, and so serves to make more acceptable DMS more tractable. Otherwise, as a result of DMS over-alerting, people may not heed safety warnings (SAE Level 0), may become upset with unexpected steering or brake adjustments (SAE Level 1), may misuse driving automation by not returning their attention when prompted (SAE Level 2), may reject and/or not be ready during control transition requests (SAE Level 3), and may miss out on important inferences of their trust/satisfaction with autonomous driving behavior (SAE Level 4-5).

Very commonly, experimental research results are caveated as depending on the situation/context. This thesis supplies ways to better know the specifics of driving scenes and driver readiness. By knowing how much eye movement is appropriate for a specific set of visible demands, the burdens of sustained driving attention and/or supervisory oversight of driving automation can be alleviated via reduction of unnecessary DMS alerts. Additionally, from the same relational/situated knowledge, driver support can be more judiciously administered and fine-tuned on an as-needed basis (e.g., adaptive back-up control) rather than in a gross sweeping way that propagates catch-22 ironies (supervising automation that purports to replace human activity) for as long as such support falls short of full-time 100% perfection and true autonomy.

Samenvatting

Probleemstelling

De Wereldgezondheidsorganisatie erkent verkeersongevallen als een volksgezondheidsepidemie met dagelijks gemiddeld meer dan 100 doden in de VS en meer dan 3.000 wereldwijd. In de zoektocht naar onderliggende oorzaken, worden menselijke fouten (met name die van onvoldoende aandacht) vaak als een hoofdschuldige geïdentificeerd. Daarom zijn de hedendaagse auto-industrie en haar voorvechters op het gebied van veiligheid erg geïnteresseerd in het bevorderen van geautomatiseerde / autonome voertuigen (AV) om zich te transformeren. Een complicatie van menselijke factoren ontstaat echter door AV-technologie op evoluerende manieren toe te passen op openbare wegen: de continue rijtaak verandert in een taak van toezicht houden en tussenkommen in geval van nood voor de automatisering van de besturing. Over het algemeen wordt verwacht dat menselijke operatoren met uitdagingen zullen worden geconfronteerd, met name voor de overgangsfasen van de SAE-niveaus van automatisering van de besturing: ongeacht of zij eindgebruikers zijn die als voltijdse opzichters (SAE-niveau 2) of als soort van veiligheidssysteem ingrijpen en op afroep werken (SAE niveau 3), of gehuurde testrijders die de ontwikkeling van autonome voertuigen overzien (SAE level 4+).

Thesis Doel en Aanpak

Binnen een groter project op het vlak van menselijke factoren bij geautomatiseerd autorijden (HFAuto, PITN-GA-2013-605817), was het doel van dit proefschrift: *'een systeem ontwikkelen dat de waakzaamheid van de bestuurder kan bewaken'*. Met een Oxford English Dictionary-definitie van *'de actie of toestand van het nauwlettend in de gaten houden voor mogelijk gevaar of moeilijkheden'*, is waakzaamheid dus betrokken bij allerlei soorten autorijden. Omdat autorijden echter geen volledige en onverdeelde aandacht vereist (ondanks tegenovergestelde gemakzuchtige aannames), treden er praktische problemen op van zodra men probeert de concepten *'voorzichtig'*, *'gevaar'* en / of *'problemen'* te operationaliseren en in het bijzonder voor systemen om de bestuurder onder toezicht te houden (DMS) waar onnodige waarschuwingen het vertrouwen van eindgebruikers, de aanvaarding en de naleving van het systeem aantasten (het zogenaamde *'cry-wolf-effect'*). Meer kennis van specifieke aandachtsbehoeften (d.w.z. in welke omstandigheden) zal naar verwachting resulteren in betere beoordelingen van de paraatheid van bestuurders hoe om te gaan met verschillende niveaus van automatisering van de besturing.

De geselecteerde benadering om te voldoen aan de doelstelling van het proefschrift was het onderzoeken van de waakzaamheid vanuit een cognitieve benadering van de systeemtechniek (ecologisch perspectief). In plaats van het concept van waakzaamheid te beperken tot een soort interne staat / eigenschap van een bestuurder, behandelde dit proefschrift de waakzaamheid als een staat / eigenschap van een systeem (d.w.z. de relatie tussen een bestuurder en een rijscène / situatie). Om te differentiëren ten opzichte van de traditionele status-quo, plaatst dit proefschrift doelbewust de kwalificatie *'situated'* om cognitie, waakzaamheid en / of DMS, enz. te beschrijven die direct rekening houdt met de huidige omstandigheden (de rijstijl) in combinatie met conventionele, op de bestuurder gerichte maatregelen.

Methoden

Onlangs heeft video-opname- en verwerkingstechnologie een exponentiële sprong voorwaarts gemaakt met behulp van gereduceerde vormfactoren en kosten. Aldus moeten camerabaseerde fysiologische en omgevingsmaatregelen (in het bijzonder het volgen van oogbewegingen en straatbeelden / segmentatie) in toenemende mate bruikbare onderzoekstoepassingsgebieden zijn ter ondersteuning van een cognitieve systeemtechnische benadering van 'situated' waakzaamheidsbewaking voor het besturen van voertuigen. Toenemende niveaus van AV-controle verminderen de handen-op-stuur en voet-op-pedaal activiteiten als informatiebronnen over het huidige gedrag van een bestuurder, dus video's (en het volgen van oogbewegingen) blijven over als bruikbare bronnen voor de beoordeling van chauffeurs. Vragen voor onderzoek / toepassing lopen uiteen van '*Wat is het aandachts- / waakzaamheidsniveau van de bestuurder?*' Naar vragen die meer lijken op '*Is de bestuurder betrokken / waakzaam genoeg voor de huidige eisen?*' Door tegelijkertijd gefilmde aspecten van de rijstijl te overwegen (en met betrekking tot de taakinhoud en -eisen). Bij het detecteren van disproporties functioneert een 'situated' stuurprogramma-monitorsysteem voor het herstellen van het nominale evenwicht tussen stuurprogramma- en scèneverereisten via verschillende soorten DMS-betrokkenheid, of het nu gaat om informatie (mededelingen, waarschuwingen, alarmsignalen enz.) En / of acties (vertraging, overgang van besturing, enz.).

Omdat nauwkeurige voorstellingen van echte werkdomeinen en ecologische beperkingen essentieel zijn voor cognitieve systeemtechnische benaderingen, heeft dit proefschrift verschillende oogmetingen van zowel nominale als afwijkende visuele besturing, onder een verscheidenheid van veeleisende en gemakkelijke rijomstandigheden zowel in het lab als in de praktijk opgetekend en gerelateerd. De huidige thesis omvat daarom een reeks methoden van hoge, middelmatige en lage betrouwbaarheid om 'situated' toepassingen van oogmeting van de bestuurder te onderzoeken in de richting van kwesties van de evaluatie van de waakzaamheid. In het proefschrift wordt theoretisch en empirisch onderzoek gebruikt in de vorm van literatuurstudie, discrete metingen van de oogbewegingen, filmopnames van straatbeelden met een bordcamera, groeperen van straatbeelden via publieksraadpleging, metingen op de weg en een rijimulator.

Resultaten en Verbindingen

Dit proefschrift bestaat uit vijf delen; het eerste deel introduceert de relevante achtergrondtheorie en het raamwerk dat ten grondslag ligt aan het proefschrift en het laatste deel bespreekt de belangrijkste conclusies. Onderdelen 2-4 richten zich op beoordelingen voor waakzaamheid van de bestuurder (Deel 2), experimenten om straatbeelden en besturdersogen te relateren (Deel 3), en de integratie van ooggebaseerde DMS met adaptieve stuurautomatisering in een rijimulator (Deel 4).

Hoofdstuk 2.1 is gericht op het karakteriseren van waakzaamheidstaken die worden toegepast in het stimuleren van onderzoek, in termen van instructies / voorwaarden, signaaltypen / snelheden en werkcomponentkenmerken in vergelijking met de klassieke literatuur op het vlak van de waakzaamheid. De beoordeling ondersteunt het belang van bewakingsdetails voor waakzaamheid (dat zijn er 18 in Tabel 2.1.1) die ontbreken voor het voorspellen / beheren van conventionele rijbewustheidsafname-situaties: specifieke consensusdefinities van conventionele

besturingssignalen, ruis en vereiste reacties. Echter, voor het toezicht op geautomatiseerd rijden, worden eigenschappen zoals de klassieke rijbewustheidsafname theorie besproken als het vergroten van de kans op problemen: temporele en ruimtelijke onzekerheid van intermitterende / zeldzame signalen die tijdkritische respons vereisen, binnen langdurige tijdsSpannes voor het uitvoeren van een taak en verhoogde eentonigheid. De conclusies uit Hoofdstuk 2.1 bevelen dus aan tot voorzichtigheid en suggereren (her-)ontwerp mogelijkheden voor het inzetten van geautomatiseerd rijden.

Hoofdstuk 2.2 stelt zes thema's voor oplossingsgebieden voor om problemen met de afname van de waakzaamheid bij het menselijke toezicht op automatisering op te lossen. Over het algemeen beschrijven de eerste drie thema's vermindingsstrategieën ofwel in een harde betekenis ofwel verschillende versies van een zachte houding: objectieve of subjectieve vereenvoudigingen van taken op het vlak van toezicht houden. De laatste drie thema's zijn gebaseerd op algemene leertheorie-paradigma's in een chronologische volgorde: gedragspsychologie, cognitivisme en ecologisch constructivisme. In het bijzonder worden de oplossingsgebieden als volgt opgesomd, benoemd en geïllustreerd. Oplossingsgebied (1): vermijd de rol van langdurig menselijk toezicht op automatisering (d.w.z. opschorten / opheffen / overslaan van automatiseringsniveaus die menselijk toezicht en een tussenkomst in geval van nood vereisen). Oplossingsgebied (2): reduceer de toezichthoudende rol langs een objectieve dimensie (dat wil zeggen, verander de duur of omvang van geautomatiseerde operaties). Oplossingsgebied (3): verminder de toezichthoudende rol langs een subjectieve dimensie (d.w.z. deel verantwoordelijkheden en / of verander de eindgebruikerervaring en -indrucken). Oplossingsgebied (4): ondersteun de superviserende rol vanuit het gedragspsychologie-paradigma (d.w.z. conditioneer het gewenste doelgedrag door middel van training en / of selectie). Oplossingsgebied (5): ondersteuning van de toezichthoudende rol vanuit het dyadische cognitivisme-paradigma (d.w.z. ontwerpen informeren ter ondersteuning van cognitieve processen en mentale modellen). Oplossingsgebied (6): ondersteuning van de toezichthoudende rol vanuit het triadische ecologische paradigma (d.w.z. ontwerpen informeren om gebruik te maken van externe omgevingscontexten en / of taakoverwegingen).

Resultaten uit Hoofdstuk 2.2 tonen aan dat onafhankelijke beoordelaars de thema's op een betrouwbare manier konden toepassen om aanbevelingen te categoriseren uit invloedrijk onderzoek naar interactie tussen mens en automatisering. De meest voorkomende oplossingsgebieden voor het probleem om de aandacht erbij te houden tijdens het toezicht houden op automatisering omvatten degene die gericht zijn op interne cognitieve toestanden, gevolgd door diegene met een breder situationeel (taak / ecologisch) perspectief.

Alles bij elkaar genomen, benadrukken de studies van Deel 2 het belang van cognitieve en situatie-afhankelijke themabenaderingen voor het beheersen van waakzaamheidsproblemen in het algemeen, maar tonen ook een gebrek aan beschikbare praktische details (dat wil zeggen, welke eigenschappen van straatbeelden en oogmetingen van de bestuurder) waarmee men zou kunnen doorgaan met het bouwen van een '*situated*' DMS. Aldus werden toegepaste oogmetingen van de bestuurder en straatbeelden uitgevoerd in Deel 3.

Hoofdstuk 3.1 presenteert een breed maar efficiënt schema voor de categorisering van de rijscène-inhoud voor aan- / afwezigheid van een functie (Bijlage 3.1.B), bijvoorbeeld type en locaties van andere weggebruikers, voertuiggedrag, zoals rijstrookwisselingen en bochten, en details van de

infrastructuur zoals wegmarkeringen, bewegwijzering, en weggromming, enz. Hoofdstuk 3.1 bevestigt relatief hoge niveaus van nauwkeurigheid en betrouwbaarheid in aantekeningen via publieksraadpleging met behulp van dat schema. Omdat externe deelnemers aan de publieksraadpleging de scènecategorisaties ongeveer tien keer sneller hebben voltooid dan conventionele intern verbonden onderzoekers, zonder degradatie van de kwaliteit van dat werk, wordt publieksraadpleging beschouwd als een interessant potentieel voor situatie-afhankelijk rijveiligheidsonderzoek. Al met al werd de meting van aspecten van rijscènes vastgelegd op een concrete en haalbare manier, wat suggereert dat gecontextualiseerde rij-informatie niet te vaag / lastig te verzamelen en vast te leggen is.

Hoofdstuk 3.2 bepaalt specifieke kenmerken van het rijscenario (d.w.z. de weggromming en het verkeer) om van belang te zijn voor de waargenomen rijprestaties en het bijbehorende gedrag, in plaats van cognitieve oogbewegingen (d.w.z. grootte van de oogsprong). Het grote aantal van aantekeningen voorziene scènesegmenten in Hoofdstuk 3.1 (~ 12.862 scènes uit ongeveer 50 verschillende rijvideo's) maakte een selectie mogelijk van stimulusmateriaal met voldoende resolutie om voorspellende regressieanalyses uit te voeren in Hoofdstuk 3.2 (dat wil zeggen, continu geschaalde onafhankelijke variabelen om continu geschaalde afhankelijke variabelen te evenaren). Een van de resulterende vergelijkingen vertegenwoordigt bijvoorbeeld de hoeveelheid waargenomen inspanning die kan worden verwacht in de aanwezigheid van specifieke hoeveelheden inhoud van de rijscène, terwijl een andere, de resulterende grootte van de oogsprong. Met name de metingen van oogbewegingen op lager niveau vertoonden sterkere (betrouwbaardere) relaties met waargenomen inspanningen en zichtbare scène inhoud (laterale / longitudinale conflicten) dan de hoger niveau weergave (en oogmetingen) aspecten van informatieopname (fixatieduur) en toegenomen cognitieve verwerking (pupilgrootte).

In Hoofdstuk 3.3 worden zowel oogbewegingen op de weg als aspecten van de rijstijl gemeten. 'Out-of-the-loop'-ogen vertoonden over het algemeen grotere excentrische bewegingsafstanden over hele reizen. Er werd echter waargenomen dat de excentrische afstanden van 'in-the-loop'-ogen periodiek stijgen en dalen met respectievelijk lage en hoge rijscène-vereisten (zoals geoperationaliseerd door stuurhoek, aantal verkeerssituaties en snelheid).

Alles bij elkaar genomen, benadrukken de studies van Deel 3 de levensvatbaarheid van het meten van relaties tussen ogen van de bestuurder en autoraces op gedragsniveau. Een toepasselijke 'situated' conclusie van DMS was dat specifieke meetbare (zichtbare) scènevereiste-kenmerken van weggromming en verkeerstelling betrouwbaar konden worden voorgesteld in pre-cognitieve oogbewegingsmetingen op laag niveau. Vervolgens voerden de studies van Deel 4 ontwerpvalidaties van een simulator proefmodel uit van verschillende integraties van real-time waakzaamheid DMS en aanstuurautomatisering.

Hoofdstuk 4.1 implementeert een real-time DMS voor het rijimulator-proefmodel en de automatisering van de rijdynamiek (dwz waarbij de automatisering dient als 'back-up' voor een bestuurder die te lang wegbleft) die verbeteringen toont in de veiligheid en de acceptatie ten opzichte van een geëmuleerd concept van de huidige functionele toewijzingen van geautomatiseerd rijden (dat wil zeggen, waarbij de automatisering zichzelf deactiveert bij het detecteren van afleiding).

Hoofdstuk 4.2 breidt het succesvolle proefmoel uit Hoofdstuk 4.1 uit. Onoplettendheidsproblemen met het toezicht op de automatisering van de auto werden bewezen (maar ook verminderd ten opzichte van een toestand waarbij één hand op het stuur werd gehouden). 'Situated' en impliciete DMS-integratieontwerpen van adaptieve reservecontrole toonden gebruikersinteractie en prestatieverbeteringen.

Alles bij elkaar genomen, benadrukken de studies van Deel 4 problemen met momenteel uitgebrachte automatiseringsontwerpen voor auto's waar mensen toezicht houden zonder aan voortdurende fysieke betrokkenheidseisen te voldoen. Het belangrijkste is dat de studies in Deel 4 de haalbaarheid van real-time ooggebaseerde DMS-integratie bevestigen met automatisering van de besturing in de richting van praktische gebruikerservaring en veiligheidsvoordelen, niet alleen bij de inzet in een directionele richting met adaptieve reserve voor de overgang van besturing, maar ook vanaf een 'situated' versie van DMS specifiek.

Conclusies, Aanbevelingen en Impact

Uit dit proefschrift kan worden geconcludeerd dat het ontwikkelen van DMS voor waakzaamheid, oogmetingen (met name van bewegingsafstanden) en scènes (met name weggrommingen en botsingsgevaren) belangrijke en relateerbare factoren zijn. Bovendien wordt geconcludeerd dat deze factoren op haalbare manieren kunnen worden verkregen voor toekomstige inspanningen op het gebied van onderzoek en ontwikkeling. Specifiek suggereren de huidige thesisonderzoeken middelen voor DMS om gericht te zijn op het beschermen en onderhouden van het lagere fundamentele niveau of de meest innerlijke lus van rij-aandacht op gedragsniveau (in plaats van interactieve impliciete cognitieve lagen en representatieve ervaringen die erbovenop aan kunnen worden toegevoegd).

Om een automatische DMS te bereiken die bijdraagt aan de transportveiligheid, moeten we menselijke intelligentie opnemen in DMS-beoordelingen van mensen in verschillende automatiseringsniveaus. Mensen zijn een adaptieve en sociale soort die 'situated' informatie en beoordelingen als gegeven beschouwt / verwacht (vooral wanneer ze als nalatig worden bekritiseerd). Met behoud van een betekenisvolle specificiteit die missers vermeidt, moeten vermeende valse alarmsignalen van eindgebruikers worden verminderd door DMS-gebruik van gedragsgerichte (visuomotorische) beoordelingen van ogen en scènefuncties samengenomen in relatie tot elkaar. Praktische aanbevelingen voor toekomstig onderzoek vallen onder twee algemene categorieën: (1) grotere betrouwbaarheid / complexiteit in rijsimulaties (bijv. meer verkeer, kruispunten en reële secundaire taken moeten de generaliseerbaarheid van naturalistische aanpassing van de bestuurder aan de eisen van de draaicirkel vergroten) en 2) grotere instrumentatietechnologie in voertuigen op de weg (bijv. betere kennis van de inhoud van de rijstijl en oogbewegingsgedrag met verbeterde meetmogelijkheden). Bovendien wordt het besturen van video-opnames aanbevolen als een groeiende onderzoeksbron die een hybride biedt van verbeterde stimulus / gedragsgetrouwheid ten opzichte van praktijkt oepassingen op de weg die ook laboratoriumniveaus van herhaalbaarheid en controle mogelijk maakt.

Van een 'situated' benadering wordt verwacht dat deze cognitieve ambiguïteit / dilemma's beter vermeidt en dient om een acceptabeler DMS beter hanteerbaar te maken. Zoniet, als gevolg van DMS-overwaarschuwing, zouden mensen mogelijkerwijs geen rekening meer houden met veiligheidswaarschuwingen (SAE Niveau 0), kunnen ze overstuur raken door onverwachte stuur- of

remaanpassingen (SAE Niveau 1), waardoor automatisering mogelijk niet wordt gereactiveerd wanneer daarom wordt gevraagd (SAE Niveau 2), kunnen ze afwijzen en / of niet gereed zijn tijdens controle-overgangsaanvragen (SAE Niveau 3), en kunnen ze belangrijke gevolgtrekkingen missen van hun vertrouwen / tevredenheid met autonoom rijgedrag (SAE Niveau 4-5).

Zeer vaak zijn experimentele onderzoeksresultaten onder voorbehoud afhankelijk van de situatie en/of context. Dit proefschrift biedt manieren om de bijzonderheden van rijtaferelen en paraatheid van de bestuurder beter te leren kennen. Door te weten hoeveel oogbeweging geschikt is voor een specifieke set van zichtbare eisen, kunnen de lasten van aanhoudende rij-aandacht en / of supervisie-toezicht op automatisering van de bestuurder worden verminderd door het aantal onnodige DMS-waarschuwingen te verminderen. Bovendien kan vanuit dezelfde relationele / 'situated' kennis ondersteuning van de bestuurder beter worden beheerd en afgestemd op een 'indien nodig' basis (bijv. adaptieve achteruitrijcontrole) in plaats van op een ruwe allesomvattende manier die hopeloze dilemma's voortbrengt (toezicht houden op automatisering die beweert menselijke activiteiten te vervangen) zolang deze ondersteuning niet volstaat voor 100% perfectie en ware autonomie.

- vertaald door "Google Translate" en Dr. ir. T. Lombaerts, Senior Aerospace Research Engineer en een goede vriend bij NASA Ames Research Center. Bedankt Thomas!

1

PART 1: Introduction

1. Background

1.1 Problems – Traffic Safety Costs

The current automotive driving domain presents formidable adverse costs for both physical and fiscal health. Nantulya and Reich (2002) have compared consequences of road traffic injuries to a worldwide societal epidemic. Within the single year of 2013, there were 32,893 motor vehicle traffic crash fatalities in the USA (NHTSA, 2018) and 1.25 million road traffic deaths across the entire world (WHO, 2018). Again within the year of 2013 in the USA, there were an approximate 1.10 fatalities per 100 million vehicle miles traveled (NHTSA, 2018) and with an estimated 2.99 trillion miles driven that year (FHA, 2018), reflects an average of about 90 people dying on the roads every single day. Beyond loss of life, other losses from car crashes can be substantial for a country's economy, including: property damage, lost earnings, medical costs, emergency services, travel delays, lost time at work, quality of life and/or legal fees. In the USA in 2010, highway accidents alone produced \$836 billion of costs, representing equivalencies of an annual expense of \$2,708 per person if spread evenly across the entire population of 308.7 million people, 5.6 percent of the \$14.96 trillion real USA Gross Domestic Product, and an estimated realized total tax payer cost of \$18 billion which approximates \$156 of additional taxes paid by every household (Blincoe et al., 2015).

1.2 Causes – Supposed Human Culprits

Human errors have been predominately blamed for vehicle traffic fatalities and accidents. The USA Department of Transportation Secretary has declared that '*the major factor in 94 percent of all fatal crashes is human error*' (NHTSA, 2017). Compared to vehicle factors and road/atmosphere conditions, drivers have been implicated in a vast majority of causes for crashes with cited problems including: inadequate surveillance, distraction, and inattention (NHTSA, 2008). Crash data from 2010 showed that 17 percent (an estimated 899,000 crashes) of all police-reported crashes involved some type of driver distraction (NHTSA, 2013). In a 50 year review of driving safety research, Lee (2008) relates that crashes are often caused by drivers failing to look '*at the right thing at the right time*' and cites supporting evidence showing that even short glances away increase crash risk (Klauer et al., 2006).

1.3 Solutions – Automated/Autonomous Vehicle Technology

The automotive industry has previously deployed advanced driver support systems (ADAS) that have saved many lives yet still see slow market uptake (Kyriakidis et al., 2015). Furthermore, the industry is also now developing automated/autonomous vehicles (AVs). Various DARPA multi-million dollar driving challenges (i.e., 2004 Grand Challenge, 2005 Grand Challenge, and 2007 Urban Challenge) (Wikipedia:DARPA Grand Challenge, 2018) have served as significant catalysts. In 2009, Google embraced winning participants from those challenges to lead and develop its own '*self-driving car project*' (Wikipedia:Waymo, 2018). Thus, the so-called '*Google Car*' became a uniquely positioned front-runner, given not only its DARPA head start, but also its Google-backed world-wide-web sphere of influence and potential to captivate audiences everywhere. Envisioned automotive AV benefits have since included aspects of increasing traffic efficiency (Van Arem et al., 2006), reducing pollution (Spieser et al., 2014), and eliminating traffic accidents and/or fatalities (Gao et al., 2014). By now, nearly every automobile manufacturer is investing in research, development and deployment of various forms of AVs. However, autonomous vehicles have also been placed along an emerging technology hype cycle (Panetta, 2017) where there are risks of

'inflated expectations' and a *'trough of disillusionment'* before a slow *'slope of enlightenment'* can be climbed towards an eventual *'plateau of productivity'*.

1.4 Complications – Continual Evolution of Imperfect AVs

AVs are continuing to evolve within and between different stages of release and development. In order to anticipate and understand potential issues of vigilance, misuse, and monitoring requirements (e.g., complacency), critical looks are required at the evolving *'state of the art'*. Concern about companies' readiness for widespread deployment of AVs (esp. while lacking a stronger regulatory leadership role from NHTSA) has been expressed by a human-automation interaction expert in a congressional testimony (Cummings, 2016). From a RAND Corporation report, Kalra & Paddock (2016) calculated that self-driving cars need to drive 275 million miles without a fatality in order to verify them to be as safe as human drivers (and sometimes hundreds of billions of miles would be needed to demonstrate their reliability). Publicly available information regarding reliability performance of AVs should reasonably be expected to constitute a critical causal factor in developing calibrated trust and end-user expectations in order to support appropriate interactions with AVs. Recently, Hancock et al. (2019) offered recommendations to address such AVs challenges:

'Two vital elements here concern calibrated operator trust and communicated transparency. For the former, design processes should seek to design explicitly for appropriate levels of trust by human occupants in light of the known reliability of the automation ... This goal is difficult, but achieving it is critical. It is difficult because we are still finding our way in understanding the contextual reliability of differing forms of automation and semiautomation offered by various manufacturers. It is critical, because if there is insufficient human trust in autonomous and semiautonomous systems, there will be both little usage and chronic misuse ...'

Despite their envisioned collective success and eventual impact (e.g., by October 2018 Waymo has logged more than 10 million miles driving in autonomous mode on public roads since 2009), even back in the sparse desert environments, or relatively controlled conditions of the urban air force base courses, the DARPA competition AVs were far from perfect. For example, in the first 2004 competition no AVs finished the 150 mile route, and instead the furthest distance achieved was only about 7.32 miles. In 2005, only 5 of the 23 AVs completed a 132 mile course. In 2007, 6 out of 11 AV finalists completed a 60 mile urban area course in the allotted 6 hour timeframe. Additionally in 2007, the contest also featured both robot collisions (with each other, pillars, and abandoned buildings) as well as robot traffic jams (Markoff, 2007).

About a decade later, the California Department of Motor Vehicles (CA-DMV) reported that there were 61 autonomous vehicle testing permit holders operating on the public roads of California (as of January 2, 2019). Consequently, CA-DMV has been evolving standardized reporting requirements for issues such vehicles are facing in terms of both disengagement and collision reports. Thus, in California, the rate of disengagement incidents for autonomous cars driven on public roads can be observed to average about one for every 716 miles (Bhuiyan, 2017a, Bhuiyan 2017b) when averaging across eight different companies testing AVs (Max: 5000 miles, Min: 0.68 miles). More formally, Favaro et al. (2018) computed cumulative disengagements as a function of cumulative reported autonomous miles and after learning effects were shown to exponentially decrease rates in the first 1 million miles, an average *'steady-state'* frequency was determined to be at around one

disengagement per around 210 miles. While *disengagements* can be caused by a multitude of reasons, range in terms of severity, and come from various sources such as the vehicle or the driver/supervisor (as detailed in Favaro et al., 2018), the *accident* rates of AVs have also been computed by Favaro et al. (2017) to be one order of magnitude worse when compared to conventionally driven vehicles '*with a mean mileage before a crash for conventional vehicles of about 500,000 miles, compared to 42,017 miles for AVs*'.

Within a climate of a technological automotive arms-race and consumer expectations, the on-road automated driving committee of the International Society of Automotive Engineers (SAE) produced a widely adopted standard J3016 in 2014 (SAE, 2014). It has been revised twice already (SAE, 2016; SAE, 2018a), and its most renowned chart another time still in December 2018 (SAE 2018b, Fig. 1.1.), to describe operational definitions to support a common language for discussion and development within the AVs community. In their words, the J3016 was issued, in part, '*to speed the delivery of an initial regulatory framework and best practices to guide manufacturers and other entities in the safe design, development, testing, and deployment of highly automated vehicles (HAVs)*' (SAE 2018b). Akin to Sheridan and Verplank's seminal (1978) '*Levels of Automation*', the SAE '*Levels of Driving Automation*' extend beyond a simplistic all-or-none notion of manual/autonomous control, by providing a graded approach that conveys a sequence of progressive steps of increased automation involvement in the dynamic driving task.

		SAE LEVEL 0	SAE LEVEL 1	SAE LEVEL 2	SAE LEVEL 3	SAE LEVEL 4	SAE LEVEL 5
What does the human in the driver's seat have to do?		You are driving whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			You are not driving when these automated driving features are engaged – even if you are seated in "the driver's seat"		
		You must constantly supervise these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	
These are driver support features				These are automated driving features			
What do these features do?	These features are limited to providing warnings and momentary assistance	These features provide steering OR brake/acceleration support to the driver	These features provide steering AND brake/acceleration support to the driver	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met	This feature can drive the vehicle under all conditions		
Example Features	<ul style="list-style-type: none"> automatic emergency braking blind spot warning lane departure warning 	<ul style="list-style-type: none"> lane centering OR adaptive cruise control 	<ul style="list-style-type: none"> lane centering AND adaptive cruise control at the same time 	<ul style="list-style-type: none"> traffic jam chauffeur 	<ul style="list-style-type: none"> local driverless taxi pedals/steering wheel may or may not be installed 	<ul style="list-style-type: none"> same as level 4, but feature can drive everywhere in all conditions 	

Figure 1.1. SAE J3016 levels of driving automation. Adopted from SAE (2018b).

A dangerous dilemma found within such an evolutionary approach regarding AVs appears in the middle levels of imperfect driving autonomy (i.e., '*automation*') which while allowing for hands and feet free operations, either requires continuous active human supervision (i.e., SAE Level 2) or readiness for automation initiated return of control to manual involvement (i.e., SAE Level 3). Banks et al. (2014) argued that incrementally increased vehicle automation (along the way to full

autonomy) may contribute to safety concerns rather than overcome them via an increased pressure put on drivers to monitor both the driving environment and the behavior of vehicle subsystems. Notably, such concerns also holds for more advanced autonomy levels (SAE level 4 and higher) which are still undergoing iterative on-road test and development with required human supervision and intervention (i.e., safety/test engineers).

Decades of research from the discipline of human factors has suggested problems and ironies in putting humans into positions where they must monitor and/or back up automated processes. The highly cited study of Norman Mackworth (1948) exposed a vigilance decrement in the performance of military personnel in simulated radar detection tasks. Hancock (1991, 2013) argues that the human operator is '*magnificently disqualified*' for a particular form of sustained attentive response and that there '*can be little doubt that human beings have been aware of the putative failings of personnel engaged in long but uneventful period on watch*'. Additionally, while vigilance problems are often regarded as a case of under-arousal associated with undemanding assignments, alternative perspectives have found the opposite to explain vigilance tasks as being highly demanding (i.e., effortful and stressful) on human mental resources (Warm et al., 2008). Parasuraman & Riley (1997) has warned that '*it has become evident that automation does not supplant human activity; rather, it changes the nature of the work that humans do, often in ways unintended*'. Likewise, Bainbridge (1983) introduces ironies where automation is used to resolve human error and humans are consequentially tasked to supervise that automation (which is not perfect)—the humans are then susceptible to further errors of manual and cognitive de-skilling that come as a result from lack of rehearsal and direct involvement.

2. Driver Monitoring Systems

While perfect AVs are not yet available to fully replace the human driver responsibility, automatic attention monitors present a reasonable solution to help mitigate consequences of inadequate surveillance problems from both the original crash causes in more traditional vehicles as well as the anticipated challenges regarding human oversight of mid-level AVs. In essence, a driver monitoring system (DMS) is concerned with detections of aberrant driver states or behavior and thus equally applicable in assessing engagement whether the observed human's driving role is that of manual control (SAE Level 0), assisted control (SAE Level 1), supervisory control (SAE Level 2) or automation-backup upon request (SAE Level 3) because all entail normative requirements for driver vigilance (e.g., readiness to respond to danger) and thus some attention to the driving environment/scene. While previous DMS could rely on measures of drivers through their hand and feet activity (e.g., steering and pedal manipulation) and consequences on vehicle motions (e.g., lateral lane position and longitudinal accelerations) these will be reduced or absent as driver inference resources as the level of driving automation is increased and driver responsibility becomes more hands- and feet- free.

In driving hands- and feet- free, the use of eye-tracking technology is in general expected to help address driver distraction problems and improve traffic safety. Camera and computation technologies have recently been progressing through reduced commodity costs (smaller, cheaper) without compromise on quality (resolution, capability). Human-centered intelligent vehicles often include video based head/eye-tracking as a major system component (Ohn-Bar & Trivedi, 2016). Hecht et al. (2019) conclude that overall, (with EEG lacking practicality and subjective measures being prone to misjudgment), '*eye tracking is the technology with the most potential*', due to its

'possibility of non-intrusive measurements and the multitude of information about the driver state', but also retains further developmental needs to increase its reliability. Furthermore, Hecht et al. (2019) suggested an apparent consensus problem result of their review that *'driver state and the different constructs lack a common definition'*.

Historically, the most common form of DMS has been focused around issues of driver underload with related terms including: *'drowsiness'*, *'sleepiness'*, *'fatigue'*, *'arousal'*, etc. (Haworth & Vulcan, 1991; Barr et al., 2009; Rau, 2005; Hanowski et al., 2008; Blanco et al., 2009; Aidman et al., 2015). However, and especially from the onset of omnipresent mobile/smartphones and growing commonality of various in-vehicle infotainment options (navigation, audio media, web applications, etc.), the use of DMS has been shifting to also include the topic of driver distraction (McGehee et al., 2007; Hickman & Hanowski., 2011).

2.1 DMS with relatively lower success

Haworth and Vulcan (1991) performed laboratory tests of various fatigue monitors in the form of eye closures from a pair eye glasses, a head nod device worn over the ear, and a reaction time measure to a red dashboard light. Upon detection of an aberrant state (eye glasses and ear-piece), or lack of timely response (dashboard light), each device produced a consequential warning in the form of an audible alarm or a loud physical buzzing. The authors reported that *'the devices showed an ability to detect fatigue in some cases but were not able to maintain alertness and thus prevent performance deterioration'*. In summary of their findings, Haworth and Vulcan (1991) stated that: *'none of the devices used resulted in fewer or shorter periods of eye closure than when no device was used'* (p.13), and *'performance after the warning signal was not markedly different to before'* (p.17).

Barr et al. (2009) performed a review of 10 different commercially available and research drowsiness detection devices that were evaluated against a set of proposed design guidelines, thus resulting in a 10 (device) x 18 (criteria) assessment table. The device meeting the highest amount of criteria only met half of the criteria set. Criteria met in common across all drowsiness detection devices included aspects of being non-invasive, operating in real-time, requiring minimum training, and not distracting from driving tasks/other safety devices. Criteria missing (i.e., requiring more data than presently available) from all devices included a minimization of missed events and false alarms, normal maintenance/replacement costs, proficiency of use, functional awareness, perceived safety benefit, intent to purchase, willingness to recommend to others, and susceptibility to behavioral adaptations.

A field operational test of a drowsy detection and warning system for heavy vehicle commercial truck operators was conducted from a partnership of the Virginia Polytechnical and State University Transportation Institute (VTTI), and the Federal Department of Transportation's Volpe Center (Rau, 2005; Hanowski et al., 2008; Blanco et al., 2009). The detection/warning system comprised of a dashboard camera that used a percentage eye closure (PERCLOS) measure to trigger visual/audio alerts to seek rest or increase alertness. General conclusions reported included that drivers in the test group were less drowsy compared to baseline, drivers with favoring opinions of systems had an increase in safety benefits, and early prototypes of the device had an overall positive impact on driver safety. However, a first set of their major research questions (over 50 were included in all)

also reported showing either no practical differences (in frequency of alerts decreasing over time) or no statistical differences as follows:

- No significant difference—impact on post-alert behavior
- No significant difference—fluence drivers to get more sleep
- No significant difference—driver achievement of better quality of sleep
- No significant difference—involved in safety critical events
- No significant difference—involved in at-fault safety-critical events
- Speculative results—fewer episodes of drowsy driving that were regarded as inconclusive due to rather large numbers of false alerts.

2.2 DMS with relatively higher success

Australian army reservists in an at-risk drowsiness population regarding on-the-job duty vehicle commutes were investigated by Aidman et al. (2015). Their system comprised of a set of worn glasses that measured blink velocity to generate continuous drowsiness scores (between 0 and 10 points with one decimal point precision) at 1-minute intervals that were displayed via a monochrome dashboard LCD along with an audio alert. Significant effects of the feedback conditions were found regarding lower average drowsiness scores, as well as reductions in peak amplitudes and durations of drowsiness scores. Subjective report results included significantly perceived differences of maintaining safer driving distances and feelings of being less drowsy.

Vehicle video recordings with external coaching from human authority figures produced significantly beneficial results with teenage novice drivers (McGehee et al., 2007) and commercial truck drivers (Hickman & Hanowski, 2011). Both studies made use of vehicle acceleration trigger events (i.e., specified g-force threshold criteria exceedances) to save both forward exterior driving scene and interior cabin facing camera footage and automatically transmit these events to parents in the case of the teenage participants and to management personnel in the case of the truckers. In either case, the incidents were reviewed with the participants by the authority figure and resulted in significant reductions in safety-related events.

In summary of the above evaluated DMS applications, what appears to be most important is favorable end-user opinion/experiences, internal and external vehicle scene/situation capture, human assessments with human review/follow-up as well as continuous assessments with interval/ratio measures. In contrast, problems and difficulties are implicated in terms of binary lights or audio beeps and challenges regarding high numbers of false alerts.

3. Theoretical Framework

Collectively, the above evaluations showed mixed results of both problems and success with different sorts of DMS. Towards the previously introduced issues of inadequate surveillance for both traditional vehicles and future AVs, a scientific underpinning to account for such differences should be useful to characterize and design future DMS. Several DMS-relevant doctoral theses have been recently published regarding the related topics of maintaining/measuring adequate visual attention in driving and for such challenges specifically as posed by automated driving. Presently relevant major take-away points can be summarized as: a combination of looking away from the

road with the occurrence of unexpected events in the driving scene is very dangerous (Victor, 2005); SAE Level 2 driving automation does not necessarily facilitate the execution of other tasks, but even the opposite which contradicts public expectations (Solis-Marcos, 2018); and physiological driver state assessment should be combined with '*data from outside the vehicle (information regarding the vehicle environment; e.g., surrounding traffic, traffic signs, and other geo-specific information'* (Van Leeuwen, 2019, p. 173).

In their textbook '*Display and Interface Design*', Bennett and Flach (2011) promote a paradigm shift inspired from and akin to the cognitive systems engineering of Norman (1986), Rasmussen et al. (1994), and Vicente (1999) as well as the ecological interface design work of Rasmussen and Vicente (1989, 1990), and Vicente and Rasmussen (1990). Therein, Bennett and Flach proposed a triadic framework to supersede the presently reigning dyadic perspective in regards to semiotics (i.e., the study of signs and symbols and their interpretation or use).

The roots of the presently reigning dyadic approach to interface design are traced to Ferdinand Saussure (1857–1913) considered by many as a principal influencer of the science of cognitive psychology that would later gain credence around the 1950s. Saussure framed the semiotic problem as that of interpretive mappings between signifiers (e.g., symbolic language) and that which is signified (e.g., mental concepts). Such a framework fits well with metaphors and goals of modern linguistics and computer science (i.e., matching symbols to concepts). In contrast, the work of Charles Sanders Peirce (1839 – 1914) framed semiotics in the context of relational links of objects and experiences within an ecological surround. Figure 1.2 compares and contrasts the dyadic and triadic models of semiotics from Saussure and Peirce respectively.

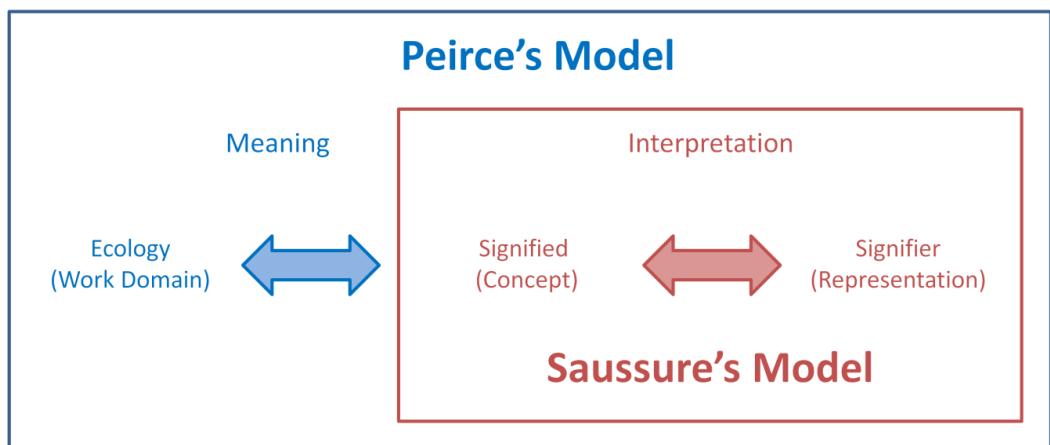


Figure 1.2. A comparison of Saussure's dyadic model of semiotics with Peirce's triadic model. Adapted from Bennet and Flach (2011), Figure 2.1, p. 18.

Beyond information processing, the triadic framework is concerned with *meaning* processing, where meaning (as understood to refer to the relation between the ecology and the signifier or representation) is the unit of interest. Such a focus, as from Bennet and Flach (2011), is in accord with James J. Gibson's notion of the direct perception of affordances that are not properties of objects or of mind but a relation of constraints/opportunities between a specific action of a specific

actor in a specific situation. For example, an affordance of walking across a sheet of ice covering a frozen lake depends on both the thickness of the ice (in reality) and the weight of the would-be walker (e.g., an ant vs. a human vs. an elephant). In other words, formulations of internal representations and resources are essentially devoid of functional meaning if not specified in relation and respects to external situations.

The situated meaning processing conceptualization of Bennet and Flach (2011), as shown in Figure 1.3, differentiates from conventional information processing approaches in several important ways. First, it is not framed in terms of processes in the head, but in terms of dynamics occurring between an actor, an information medium, and an ecology. Second, it does reflect a serial progression of processes, but an intimate coupling and parallel operation of perception and action (or control and observation). Lastly, none of the elements in Figure 1.3 is uniquely associated with either the individual or the environment – the ecology reflects the constraints scaled with respect to the organism (i.e., affordances). Bennet and Flach (2011) describe their approach of cognitive systems engineering and ecological interface design in terms of being problem-driven (as opposed to user- or technology-driven with goals of designing interfaces that (1) are tailored to specific work demands, (2) leverage the powerful perception-action skills of humans, and (3) use powerful interface technologies wisely. In other words, a principal differentiation comes from the direct treatment of situation/context which basic experimental scientists tend to want strip away as noise, but which is instead recognized as a meaningfully informative piece of the puzzle (e.g., situated cognition/action of Suchman, 1987).

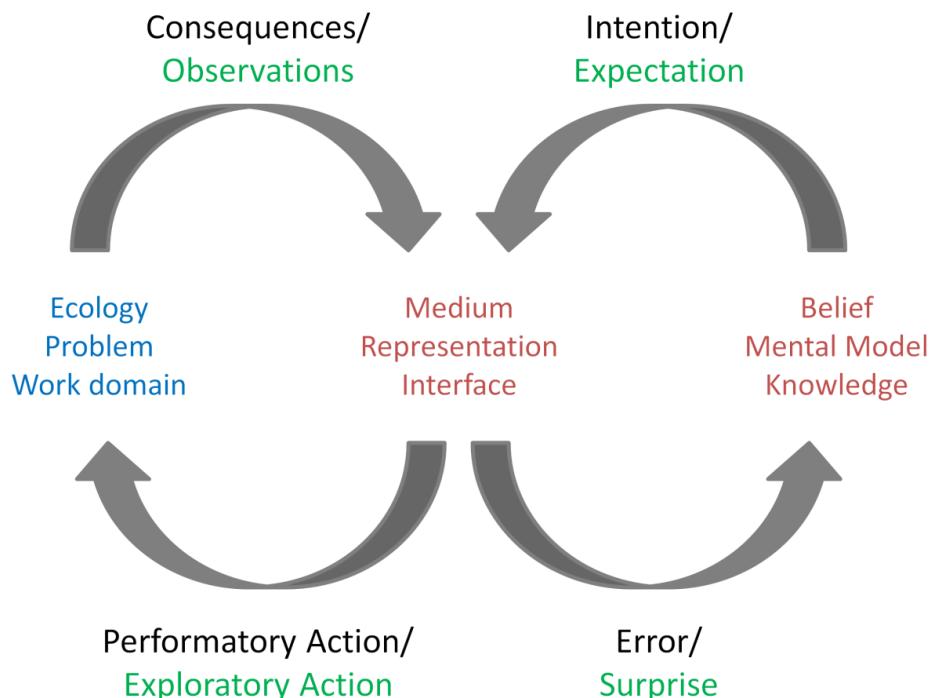


Figure 1.3. The dynamics of meaning processing involve interactions between a cognitive system (on the right) and an ecology (on the left) as mediated via an interface (in the middle). Perception and action observations are dynamically coupled in parallel with each other (terminology colored in green) and also include parallel control loops operating between consequences of actions and updates to referent-goals based from errors (terminology colored in black). Adapted from Bennet & Flach (2011), Figure 2.3, p. 32.

In terms of DMS reliability and ultimately effectiveness, these can thus be considered at different levels. From a dyadic perspective, the reliability of the DMS might be evaluated in terms of its ability to specify the monitored signal (e.g., an eye closure distance) as being present amidst measurement noise and whether those measurements might be interpreted as reflecting a construct of interest (e.g., sleepiness). From a triadic perspective, the same signals can further be evaluated in terms of meaning by consideration of the task and the environment. Using broader aspects and relational information, it is able to address ambiguities such as:

- (1) "Is the person awake enough for the present heavily trafficked urban intersections he/she is driving through?"
- (2) "Is the eye closure because the person is sleepy or because he/she is squinting under direct sunlight?"

Presumably, aspects of both the driving situation/scene and the driver change in continuous and dynamic ways and this then could be considered inconsistent with binary representations of a too-simplistic beep or buzz. More continuous value assessments would then plausibly be easier to understand, trust, and accept (cp. Aidman et al., 2015). Not only were vehicle dynamics and external scenes captured in the successful DMS intervention programs of McGehee et al (2007) and Hickman & Hanowski (2011), the assessments also included human-human discussions and elaborations of meaningfulness of the automatically triggered events in the form of reviews with an authority figure.

People expect many different kinds of benefits from different levels of AVs. However, automation benefits are easily undermined by negative user experiences and poor human-computer interactions if not designed well enough. If DMS alerts are triggered too often out-of-context (i.e., perceived false alarms), then so-called '*cry-wolf*' effects can decrease driver trust and acceptance with consequences ranging between not heeding a warning to actively seeking to defeat safety measures they deem annoying/unnecessary. For example, frustrated drivers might de-activate the DMS or not use (appropriately or at all) the driving assistance and/or automation it coincides with. Thus, the approach of this thesis was to try to understand driving vigilance issues from a situated cognition perspective of a triadic *meaning* processing rather than a dyadic *information* processing perspective (i.e., in line with the Bennett & Flach, 2011 framework). The assumption here is that a system that takes situations into account (more akin to how humans naturally do in nearly everything they do) would be more familiar and better accepted as something that is more '*smart/sophisticated*' than a closed computational model of assessment that might be too easily dismissed as '*simplistic/robotic*'. The following two example questions emphasize this subtle yet prominent difference in approach.

- (1) "How to detect and correct low levels of attention in a driver by measuring his/her eyes?"
(information processing)
- (2) "Is the observed eye behavior appropriate for the present driving task demands and what can be changed to restore a balance?"
(meaning processing)

Dyadic information processing perspectives construe meaning as an interpretation between concepts to be signified (e.g., latent internal cognitive processes) and representations that act as signifiers (e.g., physiological/behavioral measures) while then tending to avoid situations as confounding noise or difficult to interpret interactions of main effects. Unfortunately, by literally moving between contexts (across time and space) the tasks of driving (across levels of driving automation) clearly take place under a variety of demands that are hard to ignore from scientific and applied investigations that seek meaningful impact. Driving is clearly not one thing nor a task that can be cleanly separated and analyzed independent of its surrounding situations (Figure 1.4).



Figure 1.4. Different driving scene situations: high density intersection traffic in a rainy urban environment (left) and low density straight interstate freeway travel under sunny blue skies (right). Adopted from <https://youtu.be/KpGAEpm1SMs?t=43> (left), <https://youtu.be/zTB9Px6qdQ?t=33> (right).

Effective assessment of driver attention adequacy is hard to imagine without consideration of what is happening around the driver and the vehicle. Thus, it is assumed and pursued in this thesis that safety (i.e., from driver vigilance) depends not solely on measurements of latent internal driver states (arousal, attention, workload) but measurable actions (eye movements) able to be assessed relative to measurable situations (components of different driving contexts) that eyes are supposed to be adaptively working and appropriating within. The doctoral thesis of Victor (2005) reflects such motivations in the concept of '*vision-for-action*' and the Victor (2003) patent application '*System and method for monitoring and managing driver attention loads*' suggested that '*If control task intrusion is detected during secondary task glance behavior, during different road types or different demand levels, then a corresponding warning is issued*'. However, Victor (2003) did not offer further details regarding how such scene-dependencies might be practically achieved and so it is taken as a motivating research gap to which this thesis aims to contribute.

4. Thesis Aims

Across levels of driving automation, there are risks involved whenever humans become aberrant in the adequacy of their required surveillance/readiness. Automatic assessments of driver visual attention in DMS can help mitigate such risks, and eye trackers present a compelling piece of equipment that has seen massive reductions in form-factor, costs, and intrusiveness since previous generations.

This thesis was initiated in an early stage researcher (ESR) position within the Human Factors of Automated Driving (HFAuto) Initial Training Network (ITN) Seventh Framework Programme (FP7) funded by the European Commission Research Executive Agency (project number: 605817). The issued project objective was to '*answer crucial human-factors questions, such as: how should human-machine-interfaces (HMI) be designed to support transitions between automated and manual control? how can the automation understand the driver's state and intentions?...*'

The objective of this thesis was to develop a system that is able to monitor the driver's vigilance.

The selected approach to meet that objective, was to investigate vigilance from a cognitive systems engineering approach (i.e., situated cognition/ecological perspective) by including detailed considerations of driving scenes/situations within which to relate assessments of drivers/supervisors.

5. Thesis Structure

This thesis consists of five Parts; the current part (Part 1) introduces relevant background theory and the framework underlying the thesis and the last part (Part 5) discusses major conclusions drawn across the related research. In between, Parts 2-4 focus on reviews for the topic of driver vigilance (Part 2), experiments to relate driving scenes and driver eyes (Part 3), and the integration of eye-based DMS with adaptive driving automation in a driving simulator (Part 4). Several developed driving research tools are further documented and detailed alongside the research studies (as Appendices) and include: a driving scene content coding scheme (3.1.B.1), a library and interface for selecting clips with specific driving scene contents (3.1.B.2), an inexpensive apparatus for capturing on-road driving video footage (3.2.A.1), a MATLAB function for automatically clipping segments out of larger video files (3.2.A.2), a driving automation-integrated driver monitor system (4.2.A.1-2) and a programmable visual n-back GUI secondary task (4.2.A.3).

- (1) In Part 1, Introduction, a brief background picture has been painted of the human-automation interaction problems that might be expected as AV technology continues to evolve (e.g., inadequate visual attention from drivers/supervisors). Consequently, eye-based DMS was motivated as a relevant area for research and development, and in particular, a situated approach was introduced.
- (2) In Part 2, Driver Vigilance Review, literature surveys/reviews (Chapters 2.1, 2.2) are conducted to cover what has been known and done before regarding driving vigilance both before and upon the advent of AVs.
- (3) In Part 3, Driving Scenes and Driver Eyes, several experiments investigate driving scene content categorizations (Chapter 3.1) and scene-situated assessments of driver eye measures (Chapters 3.2, 3.3).
- (4) In Part 4, Adaptive Driving Automation, two driving simulator experiments were used to investigate various adaptive automation implementations of integrating an eye-based DMS with automated driving functionality (Chapters 4.1, 4.2).
- (5) In Part 5, Discussion, the results from the individual studies are re-summarized towards drawing and discussing the main conclusions across the related research at a higher level and in convergence with both broader attentional theories and recent emergent empirical evidence.

As depicted in Figure 1.5, this thesis presumes a descriptive framework model of driving monitoring systems (DMS) that serve to restore nominal balance in the face of aberrant risks. Thus, the eyes of a human supervisor within a vehicle with driving automation (or a conventional vehicle without AV technology) are presumed observable for assessment in terms of being balanced (or not) against a

given set of contextualized driving demands (traffic, signage, rules, obstacles, scenery, roadway, infrastructure, etc.). The DMS may employ alternatively heavy and/or light-handed corrections along a spectrum of automatically triggered involvement consequences (e.g., between warning information and/or driving control modification/functionality). Also Figure 1.5 is used to convey how the different publication chapters of this thesis (each with their own separate sub-goals) can approximately be represented to fit together in a comprehensive manner. This figure and summative tie-in text is re-used at the front of each journal publication chapter to serve as a re-orientation guide for the relevancy of that previously and separately published piece of research towards the overall thesis big picture.

Accurate representations of real-life work domains and ecological constraints are essential to cognitive systems engineering approaches. Thus, the present thesis includes a range of low, medium, and high fidelity methods to investigate application of driver eye movement behavior and measures towards issues of driver vigilance across levels of driving automation. Across the thesis, theoretical and empirical research was used in the form of literature survey/review, non-intrusive eye-tracking measures, dash-cam driving scene film recordings, crowdsourced content categorizations, on-road measurements and a driving simulator.

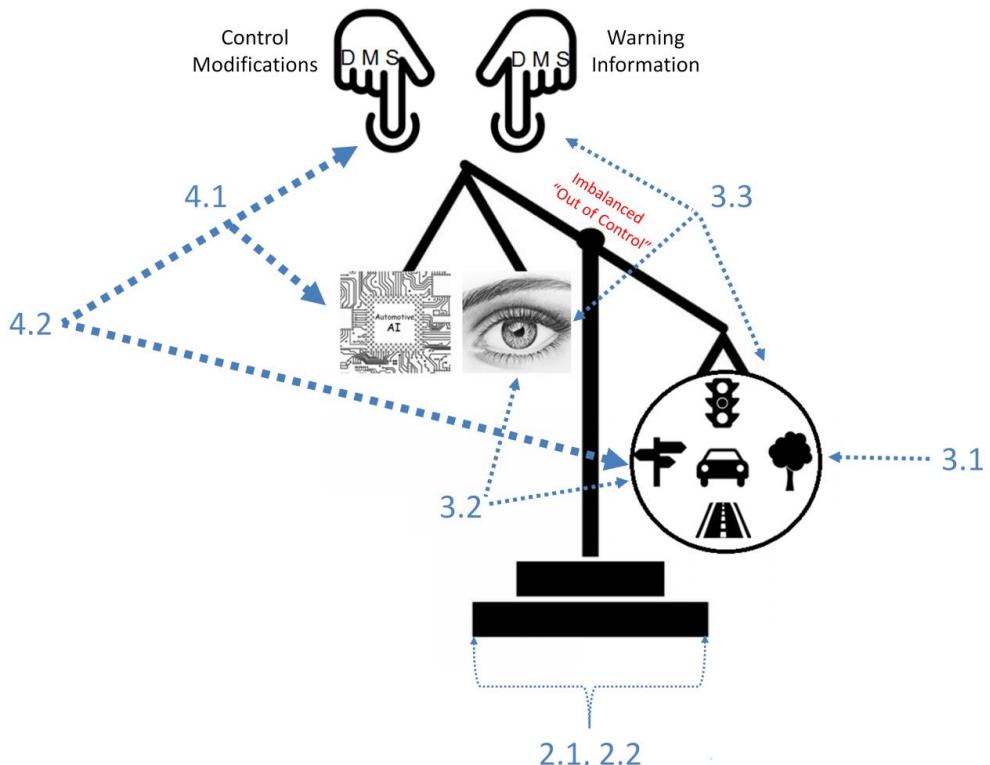


Figure 1.5. Relational mapping of publication chapters within the shared holistic coverage thesis aims to relate driver eyes, automotive vehicle automation, and driving scenes. The literature survey/review of chapters 2.1-2.2 serve as a foundation for understanding the topic of driver vigilance with and without driving automation. The investigations of chapters 3.1-3.3 orient around driving scene contents and measurable driver eye dependencies on those scene characteristics. The studies of 4.1-4.2 explore different real-time implementations and integrations of driving automation with driver monitoring systems (DMS).

References

Aidman, E., Chadunow, C., Johnson, K., & Reece, J. (2015). Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance. *Accident Analysis and Prevention*, vol. 81, pp. 8 – 13.

Bainbridge, L. (1983). Ironies of automation. *Automatica*, vol. 19(6), pp. 775–779.

Banks, V., Stanton, N., & Harvey, C. (2014). Sub-systems on the road to vehicle automation: Hands and feet free but not 'mind' free driving. *Safety Science*, 62, pp. 505–514.

Barr, L.C., Popkin, S.M., & Howarth, H.D. (2009). An evaluation of emerging driver fatigue detection measures and technologies. *U.S. Department of Transportation and Federal Motor Carrier Safety Administration*. Final Report FMCSA-RRR-09-005.

Bennett, K.B., & Flach, J.M. (2011). *Display and interface design: Subtle science, exact art*. Boca Raton, FL, USA: CRC Press.

Bhuiyan, J (2017a). Self-driving cars are mostly getting better at navigating California's public roads. *Recode*, Feb 2, 2017.
<https://www.recode.net/2017/2/2/14474800/waymo-self-driving-dmv-disengagements>

Bhuiyan, J (2017b). Uber's autonomous cars drove 20,354 miles and had to be taken over at every mile, according to documents: A first look at Uber's progress. *Recode*, March 16, 2017.
<https://www.recode.net/2017/3/16/14938116/uber-travis-kalanick-self-driving-internal-metrics-slow-progress>

Blanco, M., Bocanegra, J.L., Morgan, J.F., Fitch, G.M., Medina, A., Olson, R.L., Hanowski, R.J., Daily, B., Zimmermann, R.P., Howarth, H.D., Di Domenico, T.E., Barr, L.C., Popkin, S.M., & Green, K. (2009). Assessment of a drowsy driver warning system for heavy-vehicle drivers. *U.S. Department of Transportation, NHTSA*. Final Report DOT-HS-811-117.

Blincoe, L.J., Miller, T.R., Zaloshnja, E., & Lawrence, B.A. (2015, May). *The economic and societal impact of motor vehicle crashes, 2010. (Revised)* (Report No. DOT HS 812 013). Washington, DC: National Highway Traffic Safety Administration.

Cummings, M. (2016). Hands off: The future of self-driving cars. *Hearing before the Committee on Commerce, Science, and Transportation, United States Senate 114th Congress, second session*, Mar 15, 2016.
<https://www.govinfo.gov/content/pkg/CHRG-114shrg22428/pdf/CHRG-114shrg22428.pdf>

Favaro, F., Eurich, S., & Nader, N. (2018). Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations. *Accident Analysis and Prevention*, 110, pp. 136–148.

Favaro, F., Nader, N., Eurich, S., Tripp, M., Varadaraju, N. (2017). Examining accident reports involving autonomous vehicles in California. *PLoS ONE*, 12(9), e0184952

FHA (2018). Moving 12-Month Total Vehicle Miles Traveled [M12MTVUSM227NFWA] from U.S. Federal Highway Administration. Retrieved online Jun. 25, 2018 from FRED, Federal Reserve Bank of St. Louis,
<https://fred.stlouisfed.org/series/M12MTVUSM227NFWA>

Gao, P., Hensley, R. & Zielke, A. (2014). A roadmap to the future for the auto industry. *McKinseyQuarterly*, October.

Hancock, P.A. (1991). On operator strategic behavior. In *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 999–1007). Columbus: Ohio State University

Hancock, P.A. (2013). In search of vigilance: The problem of iatrogenically created psychological phenomena. *American Psychologist*, vol. 68(2), pp. 97–109.

Hancock, P.A., Nourbakhsh, I., & Stewart, J. (2019). On the future of transportation in an era of automated and autonomous vehicles. *Proceedings of the National Academy of Sciences of the United States of America*, doi:10.1073/pnas.1805770115, Available from:
<https://www.pnas.org/content/early/2019/01/11/1805770115.short>

Hanowski, R.J., Blanco, M., Nakata, A., Hickman, J.S., Schaudt, W.A., Fumero, M.C., Olson, R.L., Jermeland, J., Greening, M., Holbrook, G.T., Knipling, R.R., & Madison, P. (2008). The drowsy driver warning system field operational test: data collection methods. *U.S. Department of Transportation, NHTSA*. Final Report DOT-HS-810-035

Haworth, N. & Vulcan, P. (1991). *Testing of commercially available fatigue monitors*. Monash University Accident Research Centre, Melbourne.

Hecht, T., Feldhutter, A., Radlmayr, J., Nakano, Y., Miki, Y., Henle, C., & Bengler, K. (2019). A review of driver state monitoring systems in the context of automated driving. In: Bagnara S., Tartaglia R., Albolino S., Alexander T., Fujita Y. (eds) *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*. IEA 2018. *Advances in Intelligent Systems and Computing*, vol. 823. Springer, Cham.

Hickman, J., & Hanowski, R. (2011). Use of a video monitoring approach to reduce at-risk driving behaviors in commercial vehicle operations. *Transportation Research Part F*, vol. 14, pp.189–198.

Kalra, N. & Paddock, S.M. (2016). Driving to safety: how many miles of driving would it take to demonstrate

autonomous vehicle reliability? *Transportation Research Part A: Policy and Practice*, 94, pp. 182–193.

Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., & Ramsey, D.J. (2006). The impact of driver inattention on near crash/crash risk: An analysis using the 100-car naturalistic driving study data. NHTSA, Washington, DC., Technical Report no. DOT-HS-810-594.

Kyriakidis, M., van de Weijer, C., van Arem, B., & Happee, R. (2015). The deployment of advanced driver assistance systems in Europe. *22nd Intelligent Transportation Systems World Congress*. Bordeaux, France.

Lee, J.D. (2008). Fifty years of driving safety research. *Human Factors*, vol. 50(3), pp. 521–528.

McGehee, D., Raby, M., Carney, C., Lee, J., & Reyes, M. (2007). Extending parental mentoring using an event-triggered video intervention in rural teen drivers. *Journal of Safety Research*, vol. 38, pp. 215–227

Mackworth, N. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, vol. 1(1), pp. 6–21.

Markoff, J. (2007). Crashes and traffic jams in military test of robotic vehicles. *Technology. The New York Times*.

Nantulya, V.M. & Reich, M.R. (2002). The neglected epidemic: Road traffic injuries in developing countries. *British Medical Journal (BMJ)*, vol. 324, pp. 1139–1141.

NHTSA (2008). National motor vehicle crash causation survey, Report to Congress. (Report No. DOT HS 811 059). Washington, DC: National Highway Traffic Safety Administration.

NHTSA (2013). Distraction detection and mitigation through driver feedback. (Report No. DOT HS 811 547A). Washington, DC: National Highway Traffic Safety Administration.

NHTSA (2017). Automated driving systems 2.0: A vision for safety. Introductory Message from U.S. Department of Transportation Secretary, Elaine L. Chao, Guidance and Best Practices from U.S. Department of Transportation.

NHTSA (2018). Fatality Analysis Reporting System (FARS) Encyclopedia. National Highway Traffic Safety Administration. Retrieved online Jun. 25, 2018 from <https://www-fars.nhtsa.dot.gov/Main/index.aspx>

Norman, D.A. (1986). Cognitive engineering. In D.A. Norman & S.W. Draper (Eds.) *User centered system design*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ohn-Bar, E., & Trivedi, M.M. (2016). Looking at humans in the age of self-driving and highly automated vehicles. *IEEE Transactions on Intelligent Vehicles*, vol. 1(1), pp. 90–104.

Panetta, K. (2017). Top trends in the Gartner Hype Cycle for Emerging Technologies, 2017. Trends, Gartner. Retrieved online Jul. 4, 2018 from: <https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/>

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, vol. 39(2), pp. 230–253.

Rasmussen, J., Pejtersen, A.M., & Goodstein, L.P. (1994). *Cognitive systems engineering*. New York: John Wiley & Sons.

Rasmussen, J., & Vicente, K. (1989). Coping with human errors through system design: Implications for ecological interface design. *International Journal of Man-Machine Studies*, vol. 31, pp. 517–534.

Rasmussen, J., & Vicente, K. (1990). Ecological interfaces: A technological imperative in high-tech systems? *International Journal of Human-Computer Interaction*, vol. 2(2), pp. 93–111.

Rau, P.S. (2005). Drowsy driver detection and warning system for commercial vehicle drivers: Field operational test design, data analyses, and progress. *Proceedings of the 19th International Conference on Enhanced Safety of Vehicles*, Washington, DC., June 6–9.

SAE (2014). Surface vehicle information report: Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. Standard J3016. On-Road Automated Vehicles Standards Committee, SAE International. Available at: https://www.sae.org/standards/content/j3016_201401/

SAE (2016). Surface vehicle recommended practice: Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Standard J3016. On-Road Automated Vehicles Standards Committee, SAE International. Available at: https://www.sae.org/standards/content/j3016_201609/

SAE (2018a). Surface vehicle recommended practice: Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Standard J3016. On-Road Automated Vehicles Standards Committee, SAE International. Available at: https://www.sae.org/standards/content/j3016_201806/

SAE (2018b). SAE International releases updated visual chart for its “Levels of Driving Automation” standard for self-driving vehicles. SAE.org, Dec 11, 2018. Available at: <https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-%E2%80%9Clevels-of-driving-automation%E2%80%9D-standard-for-self-driving-vehicles>

Sheridan, T., & Verplank, W. (1978). Human and computer control of undersea teleoperators. MIT, Cambridge, MA. Man Machine Systems Lab.

Solis-Marcos, I. (2018). *Challenges in partially automated driving: A human factors perspective* (Doctoral dissertation, Linkoping University, Linkoping, Sweden). Available from: <http://www.diva-portal.org/smash/record.jsf?pid=diva2:1205365>

Spieser, K., Treleaven, K., Zhang, R., Fazzoli, E., Morton, D., & Pavone, M. (2014). Toward a systematic approach to the design and evaluation of automated mobility-on-demand systems: A case study in Singapore. In G. Meyer & S. Beiker, (Eds.) *Road Vehicle Automation, Lecture Notes in Mobility Series*. Springer.

Suchman, L.A. (1987). *Plans and situated actions: The problem of human-machine communication*. New York: Cambridge University Press.

Van Arem, B., Van Driel, C.J., & Visser, R. (2006). The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, pp. 429–436.

Van Leeuwen, P. (2019). *Eye movements in manual driving* (Doctoral dissertation, Delft University of Technology, Delft, the Netherlands). Available from: <https://doi.org/10.4233/uuid:308b0955-635a-4342-ac36-23e9fead164b>

Vicente, K. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. Mahwah, NJ: Lawrence Erlbaum Associates.

Vicente, K. & Rasmussen, J. (1990). The ecology of human-machine systems II: Mediating 'direct perception' in complex work domains. *Ecological Psychology*, vol. 2, pp. 207–249.

Victor, T. (2003). System and method for monitoring and managing driver attention loads, US patent application no. *US20030181822A1*, Volvo Technology AB. Available from: <https://patents.google.com/patent/US20030181822>

Victor, T. (2005). *Keeping eye and mind on the road* (Doctoral dissertation, Uppsala University, Uppsala, Sweden). Available from: <http://www.diva-portal.org/smash/record.jsf?pid=diva2:167500>

Warm, J.S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, vol. 50(3), pp. 433–441.

WHO (2018). Number of road traffic deaths. Global Health Observatory (GHO) data. World Health Organization. Retrieved online Jun. 25, 2018 from http://www.who.int/gho/road_safety/mortality/traffic_deaths_number/en/

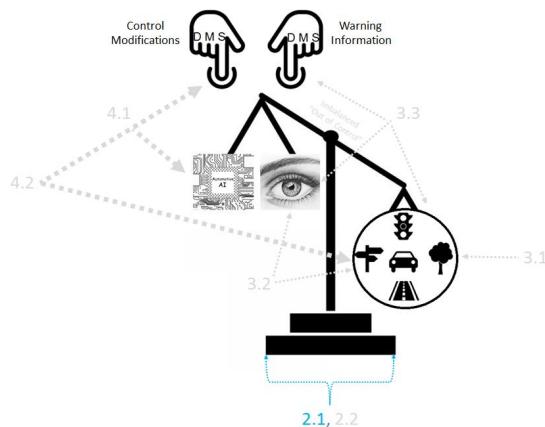
Wikipedia:DARPA Grand Challenge (2018). DARPA Grand Challenge, Wikipedia. Retrieved online Jun. 25, 2018 from: https://en.wikipedia.org/wiki/DARPA_Grand_Challenge

Wikipedia:Waymo (2018). Waymo from Wikipedia, the free encyclopedia. Retrieved online Jul. 9, 2018 from Wikipedia:Waymo/History: <https://en.wikipedia.org/wiki/Waymo>

2

PART 2: Driver Vigilance Review

Chap. 2.1) From Mackworth's Clock to the Open Road: A literature review on driver vigilance task operationalization



In regards to the overall thesis big picture, this literature review serves as a foundation for understanding the topic of driver vigilance. One of its central questions is what are drivers actually required to be vigilant of (beyond anecdotal accounts or general recommendations). As a result of the review, it appears there are not well-specified consensus answers in driving vigilance research that suggest likely vigilance decrement problems in a majority of driving tasks (i.e., little definitive overlap with classical vigilance decrement situational features). This appears to be the case as driving is recognized to be a highly variable, rather than unitary activity. However, it is also observed herein that some present designs of automated driving overlap more with classic vigilance decrement features (e.g., increasingly rare and subtle signals to which fall-back drivers must perform required responses in a consistent and time-critical manner, etc.) and thus unfortunately point towards likely vigilance problems in automated driving.

Adapted from:

Cabral, C.D.D., Happee, R., & de Winter, J.C.F. (2016). From Mackworth's clock to the open road: A literature review on driver vigilance task operationalization. *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 40, pgs. 169-189.

Abstract

Objective: This review aimed to characterize tasks applied in driving research, in terms of instructions/conditions, signal types/rates, and component features in comparison to the classic vigilance literature. **Background:** Driver state monitoring is facing increased attention with evolving vehicle automation, and real-time assessment of driver vigilance could provide widespread value across various levels (e.g., from monitoring the alertness of manual drivers to verifications of readiness in transitions of control between automated and manual driving). However, task requirement comparisons between the classic vigilance research and vigilance in car driving have not to date been systematically conducted. **Methods:** This study decomposed the highest-cited vigilance literature of each full decade since the 1940s for the situational features of the renowned vigilance decrement phenomenon originating from Mackworth (1948). A consensus set of 18 different situational features was compiled and included for example an (1) isolated (2) subject ... perceiving (3) rare (4) signals ... against (10) frequent (11) noise ... in a (17) prolonged (18) task. Next, we reviewed 69 experimental vigilance task operationalizations (i.e., required signal detection and response) within 39 publications concerned with driving vigilance. All vigilance tasks were coded as "driving vigilance tasks" or "non-driving vigilance tasks" based on the perceptual signal and response action both belonging to normal driving activity or not. Presence, absence, and unreported presence/absence of each of the 18 features was rated for each task respectively as "overlap", "contrary", and "unspecified". In conjunction, instructions/environmental conditions, signal definitions, signal rates, and summaries of the experimental vigilance tasks were extracted. **Results:** A majority of driving vigilance tasks was performed in simulators (69%) compared to on-road (28%) and watching videos (3%) along with large differences in task conditions. Participants had to maintain fixed speed/lane positions in the simulators in higher proportion (74%) than on the road (36%) where they had only to drive "normally" and/or by loose conventions like "according to the law" more often (55% versus 15%). Additionally, presence of other traffic was found more often on-road (91%) than in simulators (48%). A specification of signals to detect and react to was found present within/for driving less often (59%) than alongside/in conjunction with driving (100%). Likewise, rates of signals (i.e., frequency of signal occurrence) were reported more often for non-driving vigilance tasks (80%) than in driving vigilance tasks (21%). For driving vigilance tasks, the highest overlap was 12 of the 18 features present (67%). On average, results showed relatively low levels of classic feature overlap (36%) with high rates of unspecified feature presence (46%) for driving vigilance tasks compared to non-driving vigilance tasks with higher classic feature overlap (64%) and fewer features unspecified (13%). **Conclusion and application:** There is little overlap between the well-known and often cited vigilance decrement phenomenon and published experimental tasks of driving vigilance. Major differences were also found in the instructions/environmental conditions of simulator versus on-road experimental driving vigilance tasks. What driving vigilance practically is in the real-world thus remains a promising area for future research. We recommend that researchers apply approaches which account for more real-world driving features to better expose and address uncertainty regarding driving and vigilance.

1. Introduction

1.1. ***Timely value of vigilance operationalization for advancing driving automation***

Automobile accidents have severe costs in terms of both personal safety and financial consequence. For example, from motor vehicle crashes in the United States in 2010, there were 3.9 million non-fatal injuries, 32,999 fatalities, and economic costs totalled around \$242 billion (Blincoe, Miller, Zaloshnja, & Lawrence, 2015). Between 2005 and 2007, critical reasons for pre-crash events from a total of 5,361 analyzed crashes in a National Motor Vehicle Crash Causation Survey have been attributed to the driver in an overwhelming majority (95%) compared to vehicles (2%) and to roadway/atmospheric conditions (3%), where 48% of driver causes involved adverse driver readiness states like inadequate surveillance, distraction, inattention (e.g., daydreaming, etc.), following too closely, overcompensation, panic/freezing and/or being asleep (NHTSA, 2008). Presently researchers and industry stakeholders are rapidly progressing technological solutions within vehicles to support safer driving. Developments span a wide range of conceptual and deployed products of manufacturers and suppliers, research consortiums/initiatives, as well as information technology companies and service providers. These automotive technology developments involve a large range of categories such as driver warnings, active control assistance, and temporary or even complete relief of driving authority/responsibility. Encapsulating these developments, the German Federal Highway Institute (BASt), the United States National Highway Traffic Safety Administration (NHTSA) and the International Society of Automotive Engineers (SAE) have each produced scales for distinguishing and categorizing various levels of vehicle driving automation technology ranging from none to full (Gasser & Westhoff, 2012; NHTSA, 2013; SAE, 2014). Crucially, issues and value of knowing how to measure driver vigilance can be found throughout these aspirations and technologies in all except the absolute highest automation levels (i.e., with no human involvement at all). Definitions of vigilance are provided next before elaborating on this point.

Colloquially, the adjective 'vigilant' might only evoke images of dutiful security positions ranging from the sentinels in front of Buckingham Palace to anyone's own local neighbourhood watch program. More formally but also more broadly, Merriam-Webster and Dictionary.com respectively define vigilant as "alertly watchful especially to avoid danger" and "keenly watchful to detect danger; wary . . . ever awake and alert; sleeplessly watchful" fitting many more situations, that is, seemingly any involving purposeful watching with some adverse consequence at stake. Most well cited (and maintaining a broad coverage area), the seminal operational research definitions of vigilance classically stem from the British scientist Norman Mackworth. Within his classic WWII radar era article "The breakdown of vigilance during prolonged visual search" subjects were tasked to watch an experimental clock hand for specific sized movements (Mackworth, 1948). Mackworth first cites the usage of the term vigilance from the esteemed neurologist Sir Henry Head as "both a physiological and psychological readiness to react." Immediately after which, Mackworth then treats vigilance as "a useful word to adopt, particularly in describing a psychological readiness to perceive and respond, a process which, unlike attention, need not necessarily be consciously experienced" (Mackworth, 1948, p. 6). Thus, the present analysis follows in broadly treating vigilance tasks as any involving the ability to meet required perception and response demands.

In the context of driving, difficulties in drivers meeting required perception and response demands may be influenced from a variety of overlapping effects and mechanisms, such as widely investigated and closely related constructs of driver fatigue, driver distraction and hazard

perception. Fatigue can be characterized both as a physiological sleepy or drowsy state of a driver detectable from signature activities regarding the eyes, head, and face (Ji, Zhu, & Lan, 2004), as a psychological state of subjectively experienced disinclination to continue performing the task at hand (Brown, 1994), or relating to both physiological and psychological processes reflecting a general decreased capacity to perform (Thiffault & Bergeron, 2003). Definitions of driver distraction include not only shifts of attention away from driving stimuli/tasks (Steff & Spradlin, 2000) but also incorporate aspects of consequence (i.e., impact/effect), sources internal/external to the vehicle (i.e., activity/event/object/person) and modality types (i.e., auditory, biomechanical, cognitive, visual, or a combination) (Pettitt et al., 2005; Young & Regan, 2007). Hazard perception is a natural combination of both dangerous situations on the road ahead (Horswill & McKenna, 2004) as well as a skill developed through experience for recognizing and responding to such hazards in decreasing amounts of time (Wetton et al., 2010). Whether these constructs (and even more like workload, attention, arousal, stress, etc.) are considered independent/orthogonal, e.g., drivers may exhibit reduced vigilance (distraction) even in non-fatigued states (fully awake) and suffer performance decrements in states of both over- and underload or whether they are dependently tied, e.g. vigilance decrement as direct effect of fatigue/sleepiness is an open area of relational representation (Heikoop et al., 2015; Stanton & Young, 2000) beyond the scope of this review. However, regardless of the specific boundaries drawn by different terminology usage, such constructs (including the present topic of vigilance) all share extended consideration and coverage of both endogenous factors (i.e., emanating from within) of both physiological and psychological processes as well as exogenous factors (i.e., originating from outside).

Accurate accounts of driving demands are prerequisite to designing roles and responsibilities for various automated and/ or human driving agents. The value vigilance stands to contribute across driving and automation is detailed next by taking a step-by-step account of the NHTSA levels of vehicle automation as specific example. In the NHTSA Definitions of Levels of Vehicle Automation (NHTSA, 2013), the categories begin with Level 0 – No Automation (e.g., lane departure warning) and progress through four more levels: Level 1 – Function-Specific Automation (e.g., electronic stability control), Level 2 – Combined Function Automation (e.g., adaptive cruise control in combination with lane centering), Level 3 – Limited Self-Driving Automation (e.g., the 2012 Google car with human override), Level 4 – Full Self-Driving Automation (e.g., the 2014 Google car with no steering wheel, gas pedal, or brake pedal).

Starting with Level 0, a distinction is made that regardless of the presence/absence of various warnings (e.g., forward collision, lane departure, blind spot) or automated secondary controls (e.g., wipers, headlights, turn signals, hazard lights, etc.) the driver is in complete and sole command of the primary vehicle controls (brake, steering, throttle, and motive power) at all times and responsible for monitoring the roadway and safe operation of all vehicle controls (NHTSA, 2013). Clearly, responsibility is explicitly given to the driver for watching many aspects of both control devices and the roadway in this level and so safety checks of readiness in these duties of watching could be useful.

In Level 1, automation is function-specific (and independent in the case of multiple functions operating simultaneously) where the driver has overall control but can choose to cede limited authority over a primary control, the vehicle can automatically assume limited authority over a primary control, or provide added control in certain normal driving or crash- imminent situations; all of which occur without replacing driver vigilance and assuming driving responsibility from the

driver (NHTSA, 2013). Explicitly, vigilance is identified as a requirement of the driver not intended to be relieved from the use of the automation and hence, presumably would benefit from real-time verification that the driver is not over-relying on the automation and is sustaining appropriate levels of vigilance.

In Level 2, automation of controls can work in unison (i.e., hands off the steering wheel and foot off the pedal at the same time) however the driver is still responsible for monitoring the roadway and safe operation and expected to be available for control at all times (i.e., short notice, no advanced warning) (NHTSA, 2013). When a driver is expected for short and no- notice transitions of control, the real-time assessment of his/her readiness could be critical for safe operations to avoid startle/upset and/or loss of control.

In Level 3, the driver is no longer expected to constantly monitor the roadway while driving but instead to rely heavily on the vehicle to monitor for changes with the driver available only for occasional control and with sufficiently comfortable transition times (NHTSA, 2013). If only called upon occasionally for driving control, a driver's level of preparedness to react and respond can be expected to vary within a pre-determined allotted transition time depending on how far removed or closely tied to the driving situation the driver may or may not be.

Lastly, in Level 4 the driver is excused from an expectation of availability of control for an entire trip. While continual driver readiness to perceive and respond then is not a direct requirement, vigilance may still be useful to assess against risks of driver initiated control actions under inappropriate levels of readiness.

Generally, across any driving automation hierarchy and functional allocation framework, there may be value from accurate driver vigilance operationalization in recognizable ways. For more manual control levels, a driver might fall behind driving task demands for many reasons (e.g., falling asleep, becoming angry, day dreaming, inexperience, stimulus overload, etc.). Early detections of mismatches of driver watchfulness and preparedness to respond to events could be vital precursors to actual performance decrements and hence promote active safety through prevention rather than merely passive safety through mitigation. For partially automated situations where the driver maintains responsibility in case of automation inadequacy (or even for nominal transitions of control) and he/she is tasked to observe, the driver is expected to be ready to uptake control. Methods for actively verifying this preparedness could add value by obviating the vulnerability of merely assuming watchful readiness. Lastly, from partial and into more highly automated situations, a real-time qualification and quantification of driver vigilance can provide practical information regarding how close/far away a driver might be from the driving task demands (especially when they are allowed/encouraged to uptake additional tasks) and so can support requirements for assuring drivers back into the control loop in safe and appropriate ways. All of these aspirations for improved driving safety through none to some levels of automation entail grounding knowledge and operationalization of driving vigilance (i.e., specifically what and how drivers need to be watchful and ready for) and would be expected to be informed by established literature on human capabilities for vigilance in general.

1.2. Classic general vigilance literature

As a starting point for reconciling the above interests and values in the advancing domain of driving task responsibility evolution, the current review seeks to first look back towards classic human factors knowledge regarding the heavily researched vigilance decrement phenomenon (Mackworth, 1948), before progressing forward with future driving vigilance operationalization. We consider a summary of knowledge of the factors contributing to decreases of vigilance in general to support practical extrapolation of previously learned lessons to driving tasks specifically.

As described above, the automation of human control tasks can create problems in operational practice in addition to its intended benefits. This observation is supported by and established within the classic human factors literature. For example, Lisanne Bainbridge's seminal work "Ironies of Automation" introduces and discusses the ways in which '*automation of industrial processes may expand rather than eliminate problems with the human operator*' (Bainbridge, 1983). Specifically, she laments that within an automated system, a former operator may be recast to a monitoring role under which he is expected to take-over if things do not operate correctly. This is a problem because manual control skills that preclude against unstable or imprecise control all degrade without direct practice and use. Furthermore, cognitive strategies for appropriate control in novel or unusual situations rely on sufficient prior exposure and experience with nominal operations, and this exposure is typically remote or occluded with the provision of automated processes. Bainbridge continues by citing Mackworth stating that "we know from many 'vigilance' studies (Mackworth, 1950) that it is impossible for even a highly motivated human being to maintain effective visual attention towards a source of information on which very little happens" (Bainbridge, 1983, p. 776).

Given the potential for grave danger and adverse safety consequences, it should not be surprising that Bainbridge was not alone in these observations and interests. By the 1980s, reviews indicated that there were already at least around one thousand published reports in the literature on the topic of vigilance since WWII (Craig, 1984, Wiener, 1987). Furthermore, concerns were expressed as early as 1962 that with investigators of vigilance behaviour "spread over several continents and publishing under the sponsorship of numerous military, industrial and academic organizations, it has become a major problem to keep up with the technical literature" (Frankmann & Adams, 1962, p. 257). As a quick and current confirmatory check only of the topic's proliferation, a Google Scholar search (March, 2015) was made of titles since Mackworth in 1948 and revealed 8,140 results for vigilance and 1,540 results for its synonym sustained attention in the title only; the sum together of which (minus double-counts for appearances of both terms in the title) stood remaining at 9,652 total results. Indeed by 1987, enough material and interest had amassed on the topic of vigilance and sustained attention that a full special issue of the Journal of the Human Factors and Ergonomics Society was centralized around this topic only; an initiative itself in commemoration of the end already, of at least one entire career spent in pursuit of the same (Warm & Parasuraman, 1987). For example, the recent review of Chan (2008) provides an encompassing account of theoretical aggravators (e.g., lack of reinforcement feedback, inaccurate estimations of signal probability, irregular spatial/temporal and successive presentation of signals and events, etc.) and alleviators (e.g., increase in signal rate, self-paced tasks, greater signal intensity, etc.) of the vigilance decrement. The lessons learned in classic vigilance literature often revolve around general theoretical terminology regarding signals. To apply their solutions to vigilance assessment and decrements in driving, it is then pre-requisite to identify in driving, what exactly constitutes such signals and other relevant and potentially interacting factors or features.

1.3. Vigilance and signal stimuli concerns in the driving literature

In research publications, it is common to see driving vigilance expressed as interest/aim in many different ways. Among others, examples include: labelling driving in part or whole as some kind of vigilance task/test (Inkeri, 2010, Mets et al., 2008, Thiffault and Bergeron, 2003), vigilance as a contributing or critical factor for driving safety (CARRS-Q, 2013, Michael and Meuter, 2006, Vrignon and Rakotonirainy, 2007), driving as including/requiring large amounts of vigilance behaviour/demands (Bloomer, 1962, Boverie et al., 2008, Mackie and O'Hanlon, 1977) or driving as being comparable to/resembling a vigilance task (Atchley and Chan, 2011, Chan, 2008, Schmidt et al., 2007). Notably, consideration has also been raised to the unambiguous application of vigilance literature to specific driving scenarios like driver supervision of ACC control (Ervin, Bogard, & Fancher, 2000), the absence of regular engagements and distractions that are available on a normal highway/normal road versus in a tunnel (Jayakumar, Novak, Faber, & Bouchner, 2014) and to the relevancy of focus of vigilance problems on straight roads rather than in curves, where it is highly unlikely for someone to fall asleep (Giusti, Zocchi, & Rovetta, 2009).

It is been previously underscored that no reliable methods yet exist for defining a priori what a driver should be attending to (Hancock, Mouloua, & Senders, 2008). Instead, what (signal-processing) activities are critical for safe driving is seen as an unresolved issue in traffic safety (Regan, Hallett, & Gordon, 2011). Some insights and progress may be gained through retrospective analyses of crash and incident data. However, working backwards through reports and naturalistic driving video footage and coding some information processing activities more critical/correct than others still presents many ambiguous situations (Regan et al., 2011). One area where both ambiguity of driving signals as well as definitions of functional driving vigilance might be expected to be explicitly handled and resolved is under the highly controlled conditions and detailed documentation of experimental research and reporting.

1.4. Research aim and questions of the current literature review

Our review of vigilance tasks in driving vigilance experiments was undertaken to answer to the following questions.

- (1) What are (un)common experimental instructions and environmental conditions of driving vigilance tasks?
- (2) What are the types of signals operationalized in driving vigilance experiments?
- (3) What are the rates of those signals?
- (4) How much overlap resides between consensus features of classic vigilance tasks and experimental operationalizations of driver vigilance?
- (5) Where overlap is or is not found, what are the most common classic features present/absent?
- (6) What other circumstances (additional to the classic features) surround those tasks with the highest amount of overlap?

With answers to these questions, transportation researchers can use knowledge of driver vigilance to achieve better automation designs and hopefully greater levels of driving safety. Knowledge about the degree of overlap between classic vigilance tasks and experimental operationalizations of driver vigilance would allow us to infer when/how we can proceed in an informed manner. Degrees of contradiction and un-specification, on the other hand, would uncover gaps of knowledge to be addressed in future driving research.

2. Methods

2.1. Multi-decade consensus features of classic vigilance tasks

To utilize the findings and conclusions of prior research for applications to a new specific problem (in this case, of driving and driving automation), it is necessary to define shared components and characteristics between the prior and the current problem. A relevant first question becomes: what specific circumstances surround vigilance decrements so that we may make best use of already identified solutions? Returning to the seminal work of Mackworth (1948), the author devotes an entire section entitled “The Specific Problem” (p. 7) to introduce and emphasize the careful control of situational features surrounding the task of interest.

Thus, using Google Scholar, we retrieved the highest cited research with the terms ‘vigilance’ or ‘sustained attention’ in the title from each full decade inclusively since and that cites Mackworth in order to establish vetted consensus situational features of the now classic vigilance decrement (i.e., Davies and Parasuraman, 1982, Frankmann and Adams, 1962, Holland, 1958, Mackworth, 1948, Parasuraman, 1979, Sarter et al., 2001, Warm et al., 1996). Distal research domains outside of human factors and/or engineering psychology, such as from medicine or predator/prey animal behaviour were thus intentionally left out of scope. Common components were found to sufficiently relate a consensus in features namely involving (1) a subject/perceiver who between (2) signals/targets versus (3) noise/non-signifying events had (4) the work/task of perceiving and responding appropriately. In addition to mere presence/absence of these four feature object-nouns, 14 mutually exclusive descriptors were found to modify such objects, that is, feature modifier-adjectives. These were hence compiled in a chronologically additive manner to result in a present day composite of multi-decade theoretical features of vigilance tasks in general (Table 2.1.1).

Table 2.1.1. Present day composite of multi-decade consensus theoretical features of vigilance decrement situations as feature object-nouns and feature modifier-adjectives extracted from review of top-cited vigilance works of each full decade since Mackworth (1948).

Code	Feature
1	subject (a.k.a. participant, watcher, perceiver)
1a	<i>isolated (a.k.a. alone)</i>
2	signal (a.k.a. stimulus, target)
2a	<i>few (a.k.a. infrequent, occasional, rare)</i>
2b	<i>temporally uncertain, (a.k.a. unpredictable, probability not influenced by subject, random)</i>
2c	<i>difficult to perceive (a.k.a. small, near perceptual threshold)</i>
2d	<i>clearly perceptible when alerted (a.k.a. detectable, defined, unambiguous)</i>
2e	<i>short lasting (a.k.a. glimpse, transient)</i>
2f	<i>spatially uncertain</i>
3	noise (a.k.a. events, neutral, not meaningful, do not signify)
3a	<i>very similar to signals</i>

Code	Feature
3b	<i>frequent, (a.k.a. constantly encountered, high quantity, often)</i>
4	task (a.k.a. performance, work, assignment)
4a	<i>long duration (a.k.a. sustained, extended, prolonged, lengthy, continuous, in a series)</i>
4b	<i>lacking objective feedback of subject's own performance</i>
4c	<i>monotonous (a.k.a. same, consistent)</i>
4d	<i>successive presentations of signal and noise (a.k.a. a burden to or loading on memory)</i>
4e	<i>required response (a.k.a. action to take)</i>

Note. Feature object-nouns are in bold and feature modifier-adjective are in italics.

2.2. Search criteria, filtering, and scope reduction

Given finite resources, it would be untenable to aspire to an exhaustive review of every published driving vigilance operationalization. Instead, the goal of the present search was to gather a representative sample for detailed analysis from which to generalize. Accessing Google Scholar through Harzing's Publish or Perish scholarly citation software, a search of publications between the years 1948 and 2014 was conducted where the title had at least one word from a set of "vigilance" terms (vigilance, sustained attention, vigil, vigilant) in combination with at least one word from a set of "driving" terms (driving, driver, drivers, motorist, motorists, automobile, automobiles, car, cars, vehicle, vehicles, road, roads, motorway, motorways). Again, such a search was not engineered to return all relevant papers, but instead to ensure with greater chance that the returned sample would retain relevancy on the assumption that presence of target terms in the title connotes importance of that term to the research and hence would be a point for elaboration and description within the text.

Search results of 248, 8, 3, and 11 titles were returned respectively for "vigilance", "sustained attention", "vigil", and "vigilant" in combination with one of the "driving" terms. A total of 181 titles remained rising in frequency over the years (Fig. 2.1.1) after 89 exclusions were made from manually reading the title and/or abstract for those that were written in a language other than English (27), were duplicates within the same year (25) and across different years (14), were written about trains (9), did not actually have the search terms in the title (4), used vigilance regarding criminal theft (3), used driving as a verb of causality/influence and not locomotion (2), used road but did not involve driving (1), were about aerial vehicles (1), were about the vigilance of physicians of car accident victims (1), were about the deaths of children in trunks of cars (1), and described a macroscopic level vehicle traffic congestion/flow system (1).

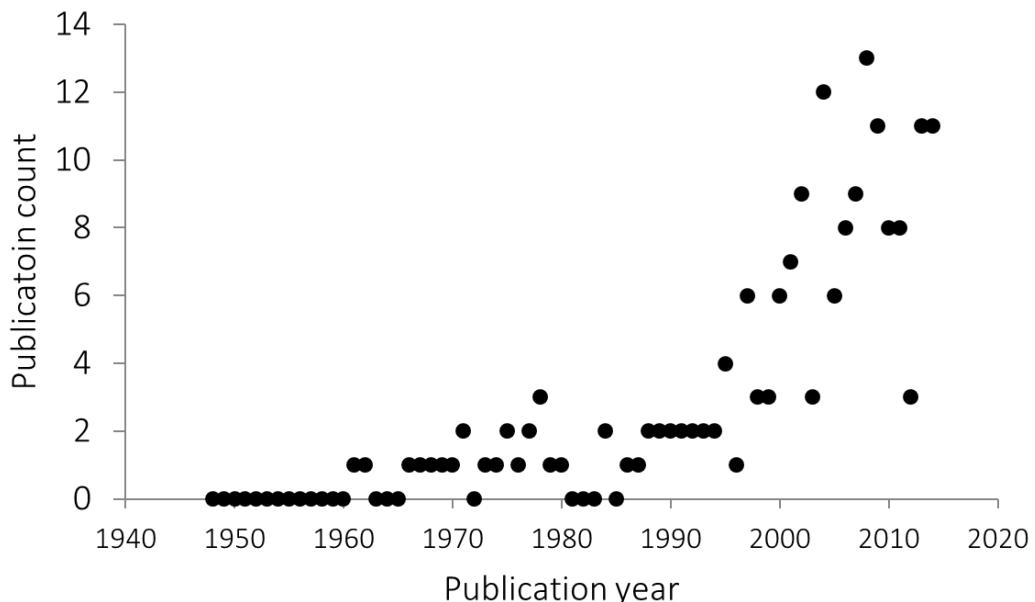


Figure 2.1.1. Publication search returns by year between 1948 and 2014 where the title had both a “vigilance” and a “driving” term in the title and did not meet exclusion criteria.

Proceeding through each of these 181 publication title returns, an approximate two-thirds majority ($n = 110$) were retrieved in full text and assessed manually for the aim of isolating experimental driving vigilance tasks. The exclusion criteria previously applied to the title/abstracts was re-applied now in greater resolution in review of full texts, and 28 more removed. Additionally, 43 were set aside that involved either algorithm/prototype validation, naturalistic observational methods or otherwise lacked explicit description of driving task experimental conditions and controlled manipulations. Where multiple experimental tasks were involved under the same title, these were expanded (30 times). Consequentially, a remaining total of 69 experimental vigilance tasks (across 39 different publications, Table 2.1.2) were eligible for analysis within the present review of sampling empirical driving vigilance task operationalizations for overlap with consensus theoretical vigilance set ups.

Table 2.1.2. Publication list of present analysis with shorthand “Ref #” index code for use in subsequent tables.

Ref #	Year	First Author	Title
1	2014	Chuang	Kinesthesia in a sustained-attention driving task
2	2014	Correa	Effects of chronotype and time of day on the vigilance decrement during simulated driving
3	2014	Jayakumar	Driver Vigilance Monitoring–Impact of the Long Tunnels
4	2014	Lin	Wireless and Wearable EEG System for Evaluating Driver Vigilance
5	2013	Amato	Effects of three therapeutic doses of codeine/paracetamol on driving performance, a psychomotor vigilance test, and subjective feelings
6	2013	Pei	Effect of Driving Duration and Work Schedules on Vigilance Level and Driving Performance of Bus Drivers
7	2013	Ruiz	Measuring the three attentional networks in a vigilance context and their relationship with driving behaviour

Ref #	Year	First Author	Title
8	2011	Atchley	Potential Benefits and Costs of Concurrent Task Engagement to Maintain Vigilance
9	2011	Schmidt	The short-term effect of verbally assessing drivers' state on vigilance indices during monotonous daytime driving
10	2010	Inkeri	Fatigue while driving in a car simulator: effects on vigilance performance and autonomic skin conductance
11	2009	Giusti	A noninvasive system for evaluating driver vigilance level examining both physiological and mechanical data
12	2009	Schmidt	Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving
13	2009	Tippin	Visual vigilance in drivers with obstructive sleep apnea syndrome
14	2009	Ueno	An analysis of saccadic eye movements and facial images for assessing vigilance levels during simulated driving
15	2008	Chan	Benefits and cost of dual-tasking in a vigilance task: A laboratory and driving simulator investigation
16	2008	Mets	Effects of Seasonal Allergic Rhinitis on Driving Ability, Memory Functioning, Sustained Attention, and Quality of Life
17	2008	Preece	Are individuals recovering from mild traumatic brain injury vigilant drivers?
18	2007	Vrignon	Impact of subjective factors on driver vigilance: a driving simulator study: in driver behaviour and training volume 3 chapter 29
19	2007	Dalton	Effects of sound types and volumes on simulated driving, vigilance tasks and heart rate
20	2007	Howard	The interactive effects of extended wakefulness and low-dose alcohol on simulated driving and vigilance
21	2007	Schmidt	Assessing driver's vigilance state during monotonous driving
22	2006	Bonnefond	Behavioural reactivation and subjective assessment of the state of vigilance—Application to simulated car driving
23	2006	Desai	Vigilance monitoring for operator safety: A simulation study on highway driving
24	2006	Michael	Sustained attention and hypovigilance: The effect of environmental monotony on continuous task performance and implications for road safety
25	2005	Lo	The impact of shift, circadian typology, and bright light exposure on sleepiness, vigilance, and driving performance in hong kong taxi drivers
26	2004	Campagne	Correlation between driving errors and vigilance level: influence of the driver's age
27	2003	Santana	Driver vigilance monitoring - new developments within the AWAKE project
28	2003	Thiffault	Monotony of road environment and driver fatigue: a simulator study
29	2002	Lucidi	The effects of sleep debt on vigilance in young drivers: an education/research project in high schools
30	2002	Roge	Alteration of the useful visual field as a function of state of vigilance in simulated car driving
31	2001	Brice	The effects of caffeine on simulated driving, subjective alertness and sustained attention
32	2001	Roge	Variations of the level of vigilance and of behavioural activities during simulated automobile driving
33	1998	O'Hanlon	Venlafaxine's effects on healthy volunteers' driving, psychomotor, and vigilance performance during 15-day fixed and incremental dosing regimens

Ref #	Year	First Author	Title
34	1995	Findley	Vigilance and automobile accidents in patients with sleep apnea or narcolepsy
35	1995	Wyon	The effects of negative ionisation on subjective symptom intensity and driver vigilance in a moving vehicle
36	1996	Wyon	The effects of moderate heat on driver vigilance in a moving vehicle
37	1978	Guillerman	Effects of carbon monoxide on performance in a vigilance task (automobile driving)
38	1976	Boadle	Vigilance and simulated night driving
39	1967	Brown	Measurement of control skills, vigilance, and performance on a subsidiary task during 12 hours of car driving

Notably, not all vigilance task operationalizations of the “driving” plus “vigilance” titled experiments were defined as belonging within nominal driving activity. For comparable and meaningful analysis, we found it necessary to further sub-divide and classify the 69 experimental vigilance tasks into mutually exclusive driving vigilance tasks ($n = 39$) versus non-driving vigilance tasks ($n = 30$). This division was made on the basis of whether both the perceptual elements to perceive and the required response actions of the task were nominally within the realm of driving or not. A representative example of each cell of this 2×2 decision matrix is given next for clarification of this classification judgement and also depicted in Table 2.1.3. Furthermore, such a division provided a baseline set of data to compare against instead of just comparing driving vigilance operationalization versus the composite consensus alone.

- (1) In Inkeri (2010) drivers were instructed to maintain a speed of 120 km/h and a central lane position and so were presumably watchful for deviations that they should correct through use of acceleration or deceleration and steering. In the present analysis, this vigilance task was classified as a driving vigilance task because both the perceptual targets and response actions reside within the notional activity of driving.
- (2) In Wyon, Wyon, and Norin (1995) driver attention was measured towards essential sources of information of varying degrees of priority within the driving task namely, indications and abnormal execution of most of the instruments, warning lamps, controls as well as auditory horn signals, noises from the engine or a near a rear wheel, and/or blue flashing (police) lights in any of the mirrors. However, the sole response required of the driver was to depress the foot switch, await an audible tone and report at leisure while holding down the foot switch and then releasing it. In the present analysis, this vigilance task was classified as a non-driving vigilance task due to the response action.
- (3) In Tippin, Sparks, and Rizzo (2009) drivers had to watch for small light targets appearing along the horizon at seven discrete locations and responded with clicking of the high beam control as soon as they detected the target. In the present analysis, this vigilance task was classified as a non-driving vigilance task due to the arbitrary perceptual targets.
- (4) In Schmidt et al. (2007), drivers were required to detect an auditory tone of 500 Hz and respond by pressing a button fitted to their right thumb. In the present analysis, this vigilance task was classified as a non-driving vigilance task due to the arbitrary nature of both perceptual target and response action.

Table 2.1.3. Experimental vigilance task division in driving research into driving vigilance tasks (++) and non-driving vigilance tasks (+, -, --) based on vigilance task perceptual elements and required response actions both belonging to notional activity of driving or not.

Experimental Vigilance Tasks	Driving response action (+)	Non-driving response action (-)
Driving percept (+)	<p>Example, Ref # 10: watch for deviations from a speed of 120 km/h and central lane position (percept, +) correct deviations with acceleration/deceleration/steering (action, +) <i>driving vigilance task (++)</i></p> <p>Full Set, Ref #s: 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39</p>	<p>Example, Ref # 35: pay attention indications and abnormal execution of most of the instruments, warning lamps, controls, as well as auditory horn signals, noises from the engine or near a rear wheel, and/or blue flashing (police) lights in any of the mirrors (percept, +) depress a footswitch and make a verbal report at leisure (action, -) <i>non-driving vigilance task (-)</i></p> <p>Full Set, Ref #s: 8, 35, 36, 38, 39</p>
Non-driving percept (-)	<p>Example, Ref # 13: watch for small light targets along horizon at seven discrete locations (percept, -) click the high beam control lever (action, +) <i>non-driving vigilance task (-+)</i></p> <p>Full Set, Ref #s: 13, 19, 25, 30, 37</p>	<p>Example, Ref # 21: detect an auditory tone of 500 hz (percept, -) press a button fitted to the right thumb (action, -) <i>non-driving vigilance task (--)</i></p> <p>Full Set, Ref #s: 2, 5, 7, 9, 10, 12, 15, 16, 18, 19, 20, 21, 24, 29, 31, 33</p>

2.3. Manual coding and annotation

Each of the 69 tasks was manually reviewed by the first author and rated against a simple ternary coding scheme for the presence, absence or unreported presence/absence of each of the 4 feature object-nouns and 14 feature modifier-adjectives seen in the multi-decade consensus circumstance composite (Table 2.1.1). Per each task, percentages of “overlap”, “contrary”, and “unspecified” were calculated by summing the number of features present (true/consistent), absent (false/contradictory), and not reported in enough detail to determine presence/absence (unreported/uncertain) respectively and dividing each sum by the total feature set size of 18 and multiplying by 100%. Furthermore, such ratings were cross-validated with 5 additional volunteer raters who redundantly and independently coded a sub sample of 4 tasks each for a total of 20 (approximately 30% of the full set of 69). A strong positive correlation was obtained between the calculated overlap percentages of these tasks rated by the additional volunteers and with those of the original rater on the same tasks ($r = .83$, Fig. 2.1.2).

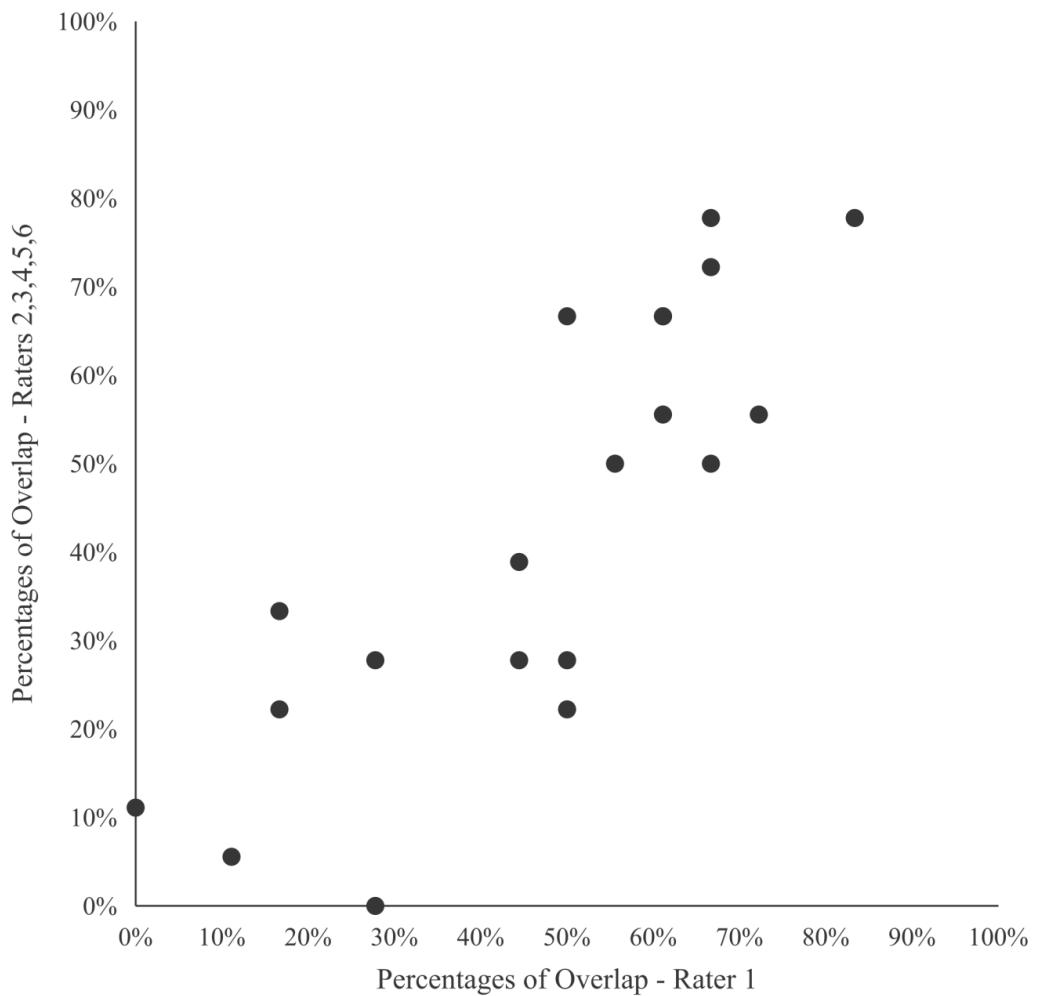


Figure 2.1.2. Correlation ($r = .83$) between the original rater and five other volunteer raters for 20 experimental vigilance tasks regarding that task's overlap with the multi-decade consensus aspects/qualifiers (Table 2.1.2).

Additional aspects of feature details and environmental conditions of the experiment were also manually reviewed and annotated for each experimental task. Specifically, the present analysis involved a qualitative identification of what the signal of interest was and a quantification of its rate of presentation scaled over one hour. Environmental conditions recorded included whether the experimental tasks took place within a simulator versus a real road; with instructions to hold fixed (or within a fixed range) a specified lateral lane position and/or longitudinal speed value; with instructions to drive “normally” and/or by some established convention/law; with the presence or absence of other vehicle traffic; and during clear visibility conditions (e.g., day time) versus deteriorated visibility (e.g., night time, fog, etc.).

3. Results

3.1. Coverage of experimental instructions and environmental conditions for driving vigilance tasks

Instructions and environmental conditions of the analyzed driving vigilance experimental tasks are presented in Table 2.1.4. Overall, a greater majority of experimental driving vigilance research was found to take place with simulators (27 of 39; 69%) compared to real-life roads (11 of 39; 28%) or use of video footage (1 of 39; 3%). Participants of the simulator studies were more often explicitly tasked with maintaining a fixed position (or a position within a fixed range) for longitudinal control (20 of 27; 74%) and/or lateral control (17 of 27; 69%) than were participants of experimental on-road tasks where lateral positions (4 of 11; 36%) and longitudinal positions (2 of 11; 18%) were mandated to be held. Contrastingly, use of more flexible instructional guidance such as “drive normally” and/or by abiding to commonly established norms, laws, and conventions was found to be higher in on road driving vigilance tasks (6 of 11; 55%) compared to simulator tasks (4 of 27; 15%). Furthermore, presence of other traffic (i.e., at least one other vehicle) was found in higher proportion in on-road (10 of 11; 91%) than in simulated tasks (13 of 27; 48%). Lastly, reporting of on-road driving vigilance tasks was found to primarily be of daytime/clear-visibility conditions (9 of 11; 82%) with lower amounts of consensus features unspecified (2 of 11; 18%). Simulator driving vigilance tasks however, were more evenly split between daytime/clear visibility (8 of 27; 30%) and night-time/reduced visibility (7 of 27; 26%) with higher amounts of consensus features unspecified (12 of 27; 44%).

Table 2.1.4. Instructions and environmental conditions of driving vigilance experimental tasks.

Ref #	Road	Sim.	Video	Lat.	Long.	Normally	Alone	Traffic	Day	Night
1		1		1	1	?	1			1
4		1		1	1	?	1			1
3		1		?	?	?		1	1	
5b		1		1	1	?	1			1
8a		1		1	1	?		1	?	?
8b		1		1	1	?		1	?	?
14		1		1	1	?	?	?	?	?
15c		1		1	1	?		1	?	?
15d		1		?	1	?		1	?	?
18b		1		1	?	?		1	1	
19b		1		?	?	?	?	?	?	?
20a		1		1	1	?		1		1
22		1		1	1	?		1		1
25b		1		1	1	?	1		1	
25c		1		1	1	?	1		1	
31b		1		?	0	1		1	?	?
34		1		1	1	?	?	?	?	?
38a		1		1	1	?		1		1
2b		1		1	1	?	1		1	
10a		1		1	1	?	1		?	?
13a		1		1	1	?		1	?	?
11		1		1	?	?	?	?	?	?
23		1		?	?	?		1	1	
26		1		?	0	1	?	?		1
30a		1		1	?	?		1		1
32		1		?	?	1	1		?	?
28		1		1	0	1	?	?	1	
6	1			?	?	1		1	1	
9b	1			?	?	1		1	1	
12a	1			?	?	1		1	1	
17		1		0	0	?		1	?	?
16a	1			1	1	?		1	1	
21a	1			?	?	?		1	1	
27	1			1	?	?		1	1	
33a	1			1	1	?		1	?	?
35a	1			?	?	1		1	1	
36a	1			?	?	1		1	1	
37a	1			1	?	?	1			1
39a	1			?	0	1		1	1	

Note. Column header indications: "Road" – driving task took place on real life road; "Sim." – driving task took place within a simulated environment; "Video" – driving task took place with videos of driving; "Lat." – subject required to maintain a fixed lane position or hold steady in a set range; "Long." – subject required to maintain a fixed speed or hold steady in a set range; "Normally" – subject asked to drive as normal/usual, by regulation, convention, law, standard, etc.; "Alone" – no other traffic present in driving task situation; "Traffic" – at least one other vehicle present in driving task situation; "Day" – daytime, clear visibility; "Night" – nighttime, fog, reduced visibility. Coding of "1" = true; "0" = false; "?" = unreported.

3.2. Types of signals

Signal type categorization of driving and non-driving vigilance experimental tasks are shown in Table 2.1.5a, Table 2.1.5b, respectively. Across the 69 vigilance tasks from the driving vigilance literature, there were 53 specified signals in total together from Table 2.1.5a, Table 2.1.5b. Fewer signals from within the driving vigilance tasks were found specified (23 signals from 39 tasks; 59%) compared to those of the non-driving vigilance tasks (30 signals from 30 tasks; 100%). The 23 identified driving vigilance signals were found to align under mutually exclusive categories in the following amounts and proportions: lateral or longitudinal deviation (12 of 23 signals; 52%), obstacles (9 of 23 signals; 39%), and light sources (2 of 23 signals; 9%). Contrastingly, the 30 identified non-driving vigilance signals were found to align under modality categories of visual (23 of 30 signals; 77%), auditory (5 of 30 signals; 17%), and multi-modal (2 of 30 signals; 7%). Further detailed descriptions follow for signal types of both the driving and non-driving vigilance tasks.

Table 2.1.5a. Signal type specification and sub-categorization of driving vigilance experimental tasks.

Ref #	Specified	Deviations			Obstacles				Light	
		Lane	Speed	Wind	Pull-out	Lead	Ped.	Hazard	?	Mirror
1	1	1								
2b	1	1	1							
4	1	1		1						
5b	1	1	1							
8b	1				1					
10a	1	1	1							
13a	1	1	1							
14	1								1	
11	1								1	
17	1							1		
16a	1	1	1							
15c	1	1	1	1						
15d	1				1					
20a	1					1				
22	1					1				
25b	1									
25c	1						1			1
30a	1					1				
28	1	1	1	1						
33a	1	1	1							
34	1							1		
37a	1	1	1							
38a	1	1								
3	?									
6	?									
9a	?									
8a	?									
12a	?									
18b	?									
19b	?									
21a	?									
23	?									
26	?									
27	?									
31b	?									
32	?									
35a	?									
36a	?									
39a	?									

Note. Column header indications: "Specified" – driving vigilance task signal definition/description specified within the text; "Lane" – a deviation from lateral lane position; "Speed" – a deviation from a longitudinal speed; "Wind" – the deviation included encouragement from an external perturbation, e.g. wind gust; "Pull-out" – a vehicle that pulls out in front of and cutting off subject vehicle; "Lead" – a vehicle the subject vehicle is following; "Ped." – a pedestrian; "Hazard" – described at general level as "a potentially dangerous traffic situation"; "?" – an obstacle without description; "Mirror" – a light source in rear view mirror; "Intersect" – a traffic intersection light. Coding of "1" = true; "?" = not specified.

Table 2.1.5b. Signal type specification and sub-categorization of non-driving vigilance experimental tasks.

Ref #	Specified	Multi-Modal	Audio	Visual				Object		Property		
				Hz	Length	Stim.	Light	Char.	Shape	Bboard	Spat. pos.	Color
2a	1								1			1
7a	1								1		1	
7b	1								1		1	
7d	1								1		1	
5a	1								1			
9a	1				1							
8c	1									1		1
10b	1						1					
13b	1						1					
12b	1				1							
16b	1							1				
15a	1							1				
15b	1							1				
18a	1				1							
30b	1								1		1	
30c	1								1		1	
37b	1						1					
38b	1						1					
39b	1						1					
19a	1						1					
19c	1								1			
21b	1				1							
20b	1					1						
24	1							1				
25a	1							1				
29	1											
31a	1							1				
33b	1						1				1	
35b	1	1										
36b	1	1										

Note. Column header indications: "Specified" – non-driving vigilance task signal definition/description specified within the text; "Multi-Modal" – multiple modalities; "Hz" – specific tone frequency; "Length" – specific tone duration; "Stim." – a visual stimulus without description; "Light" – a source of light; "Char." – an alpha, numeric, or symbolic character; "Shape" – a simple shape e.g. circle/square, etc.; "Bboard" – a billboard; "Spat. pos." – specific spatial position; "Color" – specific color. Coding of "1" = true.

Regarding the driving vigilance task signals, where the signal to detect was a deviation from maintenance of a prescribed fixed longitudinal speed (or speed range) (9 of 12 fixed position tasks; 75%) a correspondent deviation from a required fixed lateral position was also simultaneously given as a signal (12 of 12 fixed position tasks; 100%). Additionally, externally forced perturbations (e.g., lateral wind gusts) were employed in a few cases (3 of 12 tasks; 25%): once with lateral position holding only and twice with both lateral and longitudinal holding. Obstacles of different kinds were used as driving vigilance signals in the following groups and amounts from greatest to least: vehicle continuously leads ahead (3 of 9 signals; 33%); vehicle with discrete pull out or cut in ahead (2 of 9 signals; 22%); unspecified obstacles (2 of 9 signals; 22%); pedestrian leaves curb (1 of 9 signals;

11%), and examples of hazards pictured in driving scenes but not detailed in explicit description (1 of 9 signals; 11%). Lastly, in regards to the driving vigilance signals as light sources, one involved a light in a rear view mirror (1 of 2 signals; 50%) and the other a traffic signal light (1 of 2 signals, 50%).

Regarding the non-driving vigilance tasks, purely visual signals most frequently included shapes like circles or squares (8 of 23 signals; 35%); followed by sources of light (7 of 23 signals; 30%) and alpha, numeric, or symbolic characters (6 of 23 signals; 26%); with a single instance of a real life object, that is, a billboard (1 of 23 signals; 4%). Furthermore, only a handful of these included a discrimination of color in defining the signal (4 of 23 signals; 17%). For purely auditory signals, most were of a specified frequency (4 of 5 signals; 80%) with a single instance of signal definition based on duration (1 of 5 signals; 20%). Lastly, existing elements of a vehicle were rarely used and spanned multiple modalities (2 of 30 signals, 7%).

3.3. Rates of signals

Rates of signals for general and classic vigilance tasks have been already identified and discussed at length in the literature. For example by 1971, in his extensive 100+ page monograph “Vigilance: The problem of sustained attention”, road and motor vehicle traffic safety researcher Carl Stroh reviews over 35 different publications on the topic of signal frequency and concludes “when signal frequency is raised beyond a reasonable level (60–90 per hour), performance might be improved, but then it is doubtful that we are still dealing with a true vigilance situation” (Stroh, 1971, p. 8). Taking his upper bounds as the present analysis’ lower bound, signal rates less than and including 90 per hour were considered presently “few” and those greater than 90 per hour were considered absently “few” and hence not matching in terms of the infrequency of signal characteristic found within the multi-decade consensus composite (Table 2.1.1, Feature 2a). Signal rate categorization and quantifications of driving and non-driving vigilance experimental tasks are shown in Table 2.1.6a, Table 2.1.6b, respectively. A larger amount of unspecified feature adherence/contradiction was found regarding the reported rate of signal presentation for driving vigilance (31 reported signal rates from 39 tasks; 79%) than for non-driving vigilance (6 reported signal rates from 30 tasks; 20%). Where signal rates were specified in driving vigilance, slightly more than half were found to match the aforementioned frequency of “few” (5 of 8 specified signal rates; 63%). In comparison, where signal rate was more often specified in non-driving vigilance tasks, a lower proportion were found to be “few” (7 of 24 specified signal rates; 29%). A comparative depiction of proportional signal rates between the driving and non-driving vigilance tasks as well as breakouts for rates not found to be “few” is given in Fig. 2.1.3.

Table 2.1.6a. Signal presentation rates per hour of driving vigilance experimental tasks.

Ref #	Few	Rate(hr)
1	?	?
2b	?	?
4	0	133.33
5b	?	?
8b	?	?
10a	?	?
13a	?	?
14	?	?
11	1	4.8
17	?	?
16a	?	?
15c	?	?
15d	?	?
20a	?	?
22	?	?
25b	?	?
25c	?	?
30a	1	4.8
28	?	?
33a	?	?
34	?	?
37a	1	20
38a	?	?
3	?	?
6	1	60
9a	1	60
8a	?	?
12a	?	?
18b	?	?
19b	?	?
21a	?	?
23	0	360
26	?	?
27	?	?
31b	0	1560
32	?	?
35a	?	?
36a	?	?

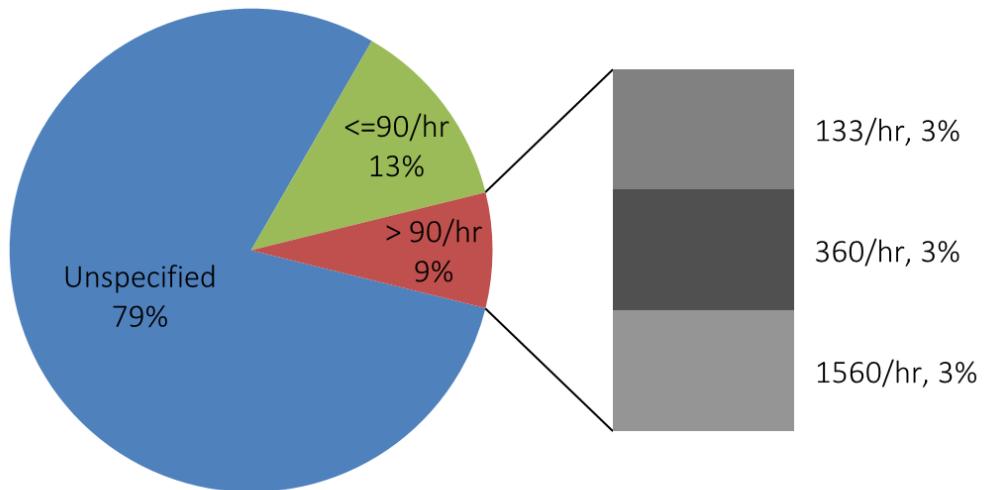
Note. Column header indications: "Few" – less than or equal to 90 presentations per hour; "Rate(hr)" = number of presentations per hour. Coding of "1" = true; "0" = false; "?" = not reported

Table 2.1.6b. Signal presentation rates per hour of non-driving vigilance experimental tasks.

Ref #	Few	Rate(hr)
2a	0	600
7a	0	219.43
7b	0	219.43
7d	0	219.43
5a	0	360
9a	0	102.86
8c	1	12.24
10b	?	?
13b	1	60
12b	0	102.86
16b	?	?
15a	0	348.84
15b	0	120
18a	?	?
30b	1	34.47
30c	0	159.6
37b	?	?
38b	1	30
39b	1	20
19a	1	60
19c	0	540
21b	0	120
20b	0	600
24	0	345
25a	0	200
29	0	144
31a	0	480
33b	1	40
35b	?	?

Note. Column header indications: "Few" – less than or equal to 90 presentations per hour; "Rate(hr)" = number of presentations per hour. Coding of "1" = true; "0" = false; "?" = not reported

Driving Vigilance Tasks



Non-Driving Vigilance Tasks

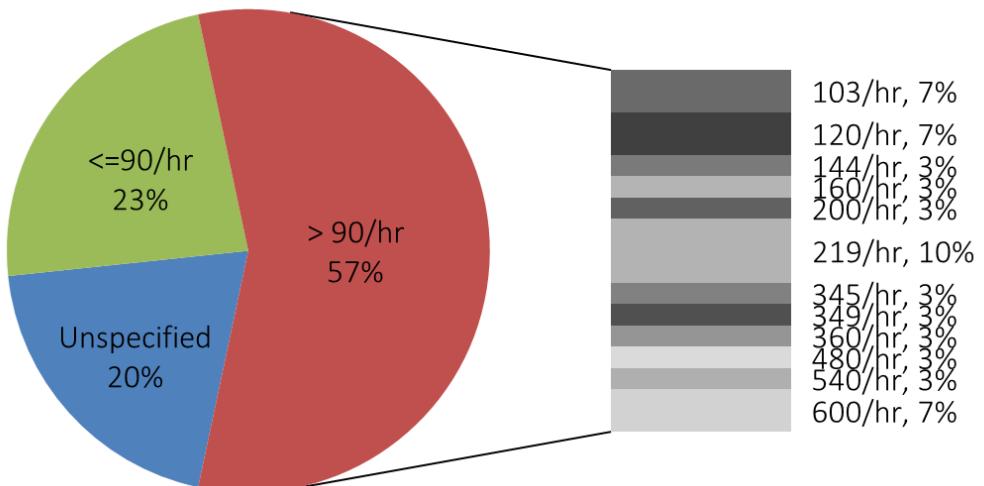
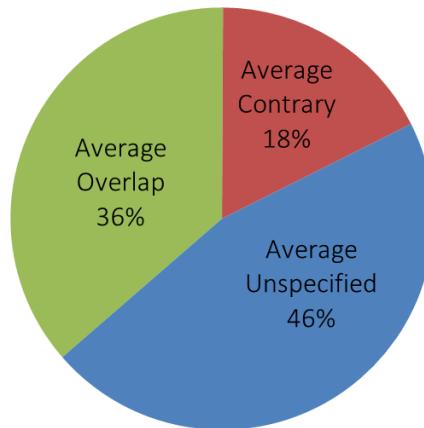


Figure 2.1.3. Specification of signal rates in driving vigilance versus non-driving vigilance tasks and stacked bar delineation for proportions of specific rates when in excess of “true vigilance situations” (i.e., $>90/\text{hr}$) (Stroh, 1971, p. 8).

3.4. Percentages of overlap with classic/general vigilance tasks

An overlap percentage was computed from the number of the multi-decade features (Table 2.1.1) that were present within each of every vigilance task operationalization. On average, less than half of the consensus features were found to be present overall ($M = 48\%$, $n = 69$) with an average amount of unspecified features of 32%. Averaging separately, however, revealed a lower average overlap for the driving vigilance tasks ($m = 36\%$, $n = 39$) with higher amounts unspecified ($M = 46\%$) compared to the non-driving tasks ($M = 64\%$, $n = 30$) with lower amounts unspecified ($M = 13\%$) as seen in Fig. 2.1.4.

Driving Vigilance Tasks



Non Driving Vigilance Tasks

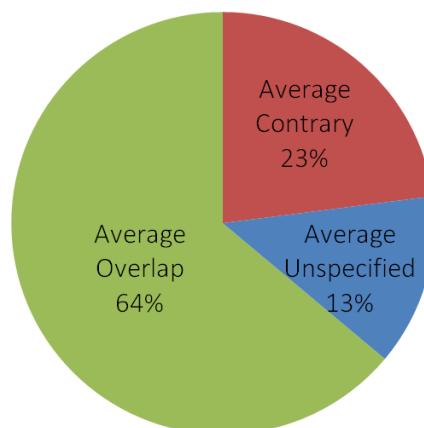


Figure 2.1.4. Averages of classic consensus feature overlapping presence, contrary absence, and unspecified presence/absence for driving vigilance tasks ($n = 39$) versus non-driving vigilance tasks ($n = 30$).

3.5. Most common features of overlap, contrary, and unspecified

For both the driving and non-driving vigilance tasks and each multi-decade consensus feature (Table 2.1.1), separate sums of the ratings of overlap, contrary, and unspecified were computed (Table 2.1.7) to determine what of classic vigilance tasks were most held in common, in contradiction or in uncertain terms. In the driving vigilance tasks, the most common feature overlapping with the classic features was that regarding a lengthy duration (i.e., half an hour or longer, Feature 4a) (28 of 39; 72%). In the non-driving vigilance tasks the most common feature in overlap was a tie between the detection of a signal (Feature 2) and the requirement of making a specified response (Feature 4e) (30 of 30; 100%). Regarding contrary features, the most common feature absent and in contradiction for the driving vigilance tasks was the successive presentation of signal and noise (i.e., a burden of memory of the distinction between these provided their non-simultaneous/overlapping occurrences, Feature 4d) (20 of 39; 51%). For the non-driving vigilance tasks the most common feature absent and in contradiction was a tie between the signals being few in frequency (i.e., <90 per hour, Feature 2a) and the signals occurring in spatially uncertain locations (Feature 2f) (17 of 30; 57%). Lastly for unspecified feature presence/absence, the provision of objective feedback of a subject's own task performance (Feature 4b) was the feature most often rated as unspecified in both driving vigilance (35 of 39; 90%) and non-driving vigilance tasks (20 of 30; 67%).

Table 2.1.7. Counts of classic vigilance features (Table 2.1.1) for driving and non-driving vigilance experimental tasks

Feature Code	"1" Driving (39 tasks)	"1" Non-Driving (30 tasks)	"0" Driving (39 tasks)	"0" Non-driving (30 tasks)	"?" Driving (39 tasks)	"?" Non-driving (30 tasks)
1	25	27	0	3	14	0
1a	14	10	14	9	11	11
2	23	30	0	0	16	0
2a	5	7	3	17	31	6
2b	14	29	9	0	16	1
2c	0	6	18	15	21	9
2d	15	26	3	1	21	3
2e	8	21	13	0	18	9
2f	3	10	16	17	20	3
3	17	18	4	11	18	1
3a	17	18	4	11	18	1
3b	14	17	6	11	19	2
4	25	29	1	0	13	1
4a	28	17	8	13	3	0
4b	2	8	2	2	35	20
4c	25	27	2	0	12	3
4d	1	15	20	14	18	1
4e	20	30	0	0	19	0

Note. Column header indications: "1" = feature presence; "0" = feature absence; "?" = not reported feature presence/absence. Feature counts exceeding half of the full set of tasks are in bold as "common" and the highest count is in bold and in italics as "most common".

3.6. Task summaries of the highest amount of overlap

Finally, additional task summaries of those experimental driving vigilance tasks with the highest amount of overlap with the classic consensus general vigilance features are next presented in a four-way tie of 67% overlap each. First, with the lowest amount of unspecified features (6%), drivers were asked to immediately steer back to the center of the original lane once perturbed by random forced departure events while on a simulated night time roadway with no other traffic and cruising at a constant speed of 100 km/h (Lin et al., 2014). Second, with the next lowest unspecified consensus features of 11%, participants steered towards randomly left/right deviating red tail lights projected at a constant distance ahead “as if following it along a country road at night” while seated in a stationary car cabin but with a forward projection of a moving road of random dot patterns on an extended table surface ahead of their cabin, whose progression was coupled to the input of their accelerator pedal (Boadle, 1976, p. 220). Lastly, in two separately coded driving vigilance tasks from the same publication (Lo, 2005) and both with 17% of the consensus features unspecified, participants had to step on a brake pedal as response to encountering either a pedestrian stepping away from the sidewalk into the driving lane or a traffic light that changed from green to red. These participants were occupational taxi drivers who performed the test while seated in their own stationary real-life taxi with a 15” laptop displaying a simulated 80 km/h flowing view of a monotonous road lacking any other traffic or lateral control.

4. Discussion

This review aimed to characterize experimental driving vigilance tasks in terms of common instructions/conditions, signal types/rates, and component features for comparison to the classic vigilance literature. From sampling experimental literature principally concerning both driving and vigilance, we found task operationalizations that were not highly similar with the full set of multi-decade consensus situational features surrounding the vigilance decrement. The overall results support critical (re)evaluation of driving tasks as being construed as vigilance tasks in the classic sense.

4.1. Coverage of experimental instructions and environmental conditions for driving vigilance tasks

Our results revealed large and informative differences between the common instruction/conditions used in experimental driving vigilance research, especially along the dimension between the use of simulators or real roads. First, and perhaps unsurprisingly, simulator studies were about twice as common as real-world settings. Furthermore, simulator studies were found to more commonly restrict the driving task into maintaining a specific speed and lane position and hence driving vigilance arises as the perception and response to deviations from such mandates. When operationalized on real roads, drivers were more often flexibly tasked with only general adherence to legal/social conventions for driving. The driving vigilance here, might then be differently construed as the perception and response to deviations from safety or normality. Additionally, a large component of driving safety can reasonably be expected to include the presence/absence of other vehicles, which was about twice as commonly available in the real-world versus the simulator studies. Real-world studies, however, were seen to more commonly be restricted to conditions of near perfect visibility compared to simulator studies which more evenly exposed driving participants to both day/clear and night/fog environment.

4.2. Types of signals

Similar challenges for the topic of driving vigilance were found from the analysis of signal types in experimental driving vigilance tasks. First, explicit descriptions of the driving specific signals a participant should be ready to perceive and respond to with a driving action were not found in over a third of what were coded as driving vigilance tasks. This shared difficulty alone suggests a decomposition of driving and assessment of driving vigilance to be potentially problematic. Potential solutions may include direct manipulations of instructions and/or stricter documentation of the specific instructions given to participants of driving vigilance experiments along with the avoidance of instructions which may be susceptible to generalities/assumptions such as to drive “safely”, “as you normally would”, “according to the law”, etc. Furthermore, clear consensus was not found between driving vigilance signal operationalizations, with a relatively even split between obstacles and speed/lane deviations and with only a few light signal sources. Unfortunately, as discussed earlier, deviations from prescribed speed or lateral positions might not be as realistic a concern of driving vigilance as the perception and avoidance of obstacles (i.e., especially other traffic). Additionally, a relative scarcity of light source signals (i.e., 2 of 23 driving vigilance signals) seems problematically disproportionate, given a large prevalence of visual light signals in real-world driving (e.g., intersection lights, caution lamps, turn signals, headlights, etc.) as well as in automated warnings/indicators (e.g., dashboard, heads up displays, etc.). Considering the possible modalities all of these driving vigilance signals might manifest through (as in the non-driving vigilance signals), additionally suggests a potential mismatch of focus. At present, a gap can be seen surrounding the use of real-life and multi-modal types of signals for experimental driving vigilance assessment and investigation.

4.3. Rates of signals

More problems for an informed identification and alleviation of vigilance decrements were found in the lack of reported signal rate/frequencies when describing the driving task specific signals and responses. This same level of unspecified signal rates (79%) was not evidenced in non-driving vigilance tasks (20%) and suggests the least difficulty in reporting, and possibly even a gap in knowledge or approach regarding frequencies of driving vigilance signals. While more than half of specified signal rates in the driving vigilance tasks were indeed within the range of a “true vigilance situation” (Stroh, 1971, p. 8) these are at a minority against the disproportionate unspecified of the majority. Thus for the accurate prediction and alleviation of vigilance decrements, the present review reveals an unfortunate lack of articulation of a presumably prudent direct consideration/exploration of exactly how rare and/or how much influence drivers might have on the signals they must respond to while driving. Generally, whenever signal-response approaches are used, it is recommended to include precise documentation of the rate of signal presentation and especially for investigations of vigilance to also include stipulations surrounding any influences a participant might have on that rate or on its being predictable for the participant.

4.4. Percentages of overlap with classic/general vigilance tasks

All the vigilance tasks of the analysis averaged together showed a weak overlap with the multi-decade consensus vigilance theory situational features (less than half on average), thus suggesting some misalignment of operationalization between theory and practice. Splitting this overlap comparison revealed less theoretical overlap for driving vigilance tasks (36%) versus non-driving vigilance tasks (64%). Additionally, the unspecified consensus aspect/qualifier presence or absence was higher for the driving vigilance tasks (46%) and lower for the non-driving vigilance tasks (13%).

In the least, it is evident then and convergent with the prior results of the present analysis, that describing driving as a vigilance task and tackling its potential for detrimental vigilance performance is not straightforward and the lessons learned thus far from vigilance theory therefore might not be readily applied. Extensions of classical definitions of vigilance to situated definitions of driving vigilance, especially pertaining to anticipated or requisite characteristics (e.g., features) may provide a way forward. Moreover, such definitions might attempt to identify features that are essential throughout all driving vs. specific to particular driving contexts or scenarios.

4.5. Most common features of overlap, contrary, and unspecified

The present analysis extended beyond the identification of a lack of consensus overlap (i.e., between shared features of classic vigilance circumstances and experimental driving vigilance operationalization), to help reveal why this might be the case. Perhaps unsurprisingly, driving vigilance tasks and non-driving vigilance tasks had similar overlap with consensus vigilance features regarding the presence of a perceiver tasked to respond to signals over a prolonged period in a consistent/unchanging standard of performance. More informatively, however, the current analysis showed features that are not commonly reported for driving vigilance tasks but which are commonly reported in non-driving vigilance tasks. These features of large quantities of non-meaningful noise events which are highly similar to target signals where the target signals themselves are not predictable and not subject to any driver influence on the probability or duration of occurrence are lacking specification in driving vigilance operationalization. Such a lacking presents direct challenges of practically matching driving vigilance problems to general classic vigilance theory. Furthermore, the successive and memory burdening presentation of signals separate from noise was found to be absent in more than half of the driving vigilance tasks where instead signals emerged from or simultaneously overlapped with their noise (e.g., a pedestrian stepping away from a curb, or lateral heading drifting away from lane center, etc.).

4.5.1. Task summaries of the highest amount of overlap

Those few tasks with the highest amount of consensus feature overlap may shed light on circumstances research could focus on for safeguarding against classic vigilance decrements. In summary of the cases with an approximate two-thirds overlap with consensus classic vigilance circumstances, decrements of vigilance might be predicted for drivers alone at night attempting to follow precise lateral positions at constant speeds, in performing correct braking responses to red traffic signal lights and errant pedestrians, or in other conceivability similar circumstances. As an example of applied vigilance solutions then, deviations from a prescribed lane center could be made more salient by auditory and visual alerting with Lane Departure Warnings. In addition to increasing the predictable/regular occurrence of encounter of pedestrians and/or traffic lights (e.g., crosswalks, intersections, etc.) such signals might be highlighted or emphasized by advanced recognition software such as with heads-up displays. However, two-thirds (while the highest found) is by definition only partial overlap and those elements missing might also be the ones crucial to or interactive with other aspects for performance in that specific situation. Until these are better understood from additional research and investigation, the driver vigilance support solutions may prove inadequate at best and inappropriately applied at worst.

4.5.2. Highest amount of overlap in highly automated driving?

Decrements and problems of vigilance may be expected to arise in future driver assistance and automated driving systems to the extent that circumstances of their use cases might resemble the

classic vigilance situational feature set. While the driving tasks of the present analysis did not often explicitly identify themselves as operating within an automated driving paradigm, some task conditions did automate lateral and/or longitudinal control in their experimental methods and so could be seen as reflecting a NHTSA Level of Vehicle Automation 1 and/or 2. Moreover, a body of driving vigilance concerns are emerging from BASt- and NHTSA-like definitions of automated driving where a driver is required to respond to an automated system take-over request provided no/short notice and/or a pre-established length of time (Gasser and Westhoff, 2012, NHTSA, 2013). While initially out of the scope of the present analysis because the title did not use the terminology of “vigilance” or “sustained attention” in its title, the take-over request (TOR) automated driving simulator experiment of Gold, Dambock, Lorenz, and Bengler (2013) maintains relevance to the present discussion as many of its theoretical and experimental task features could be considered in overlap with the classic vigilance feature set.

In Gold et al. (2013), subjects were tasked with a pre-occupying secondary task while the car drove itself until an auditory and visual alert prompted them to take-over to avoid an accident ahead of them either through braking or swerving to another lane. In their methods, 50% of the set of features of classic vigilance tasks are present with a subject watching/listening for an infrequent, temporally uncertain, unambiguous, time-critical signal that they must perform a required response to in a consistent/routine manner. However, a much higher overlap around 83% (and highest yet of any of the tasks of the present analysis) is conceivable for TORs when adding to the specific reported methods of Gold et al. (2013) features likely within TOR in general. These additional features might include an isolated driver required to respond during prolonged periods of inactivity to imperfect automation through which the driver must make asynchronous discriminations between noise (i.e., false alarm/missed events) that is highly similar to valid signals. The classic vigilance decrement features of time criticality (i.e., short lasting signals) and lack of feedback on driving response in TOR, while respectively present and unspecified present/absent in Gold et al. (2013), however should not and does not necessarily hold true in all future real-world TOR implementations. Further research and investigation is thus seen as especially needed in regards to the specific potential for decrements of vigilance provided higher levels of driving automation surrounding the situational features entailed by design, implementation and actual driver use.

5. Summary and Limitations

From reviewing experimental driving vigilance task operationalizations, the results of the present analysis have shown the topic to be of great concern but a challenge for specific consensus definition and treatment. The results are by strict definition limited to the narrow selection of literature from specific inclusion/exclusion criteria, yet may generalize beyond the use of “vigilance”/“sustained attention” and “driving” in the title. The general results of uncertainty surrounding driving vigilance operationalization might also be considered an artifact of the feature set and coding schemes undertaken. However, the marked differences observable from the non-driving vigilance tasks using these same methods serve to provide relative confirmation. Moreover subjectively, the same difficulty of complexity and articulation in driving vigilance can be appreciated merely from asking oneself which and to what extent any of the circumstances described above may or not be present when people actually drive in normal day-to-day situations.

6. Directions for Future Research

Concerns in the literature over the real-world applicability of findings from laboratory/simulator vigilance experimental tasks span multiple decades of criticism and review (Kibler, 1965, Craig, 1984, Mackie, 1984, Wiener, 1987) to the near present day (Hancock, 2013a) and are equally shared by driving safety researchers seeking theoretical transfer (Rosenbloom & Wolf, 2001).

Raising such discussion and concern in the transportation research literature can help protect against prohibitions regarding driving task requirements (i.e. perceptual targets and response actions) while these requirements are still uncertain and supports the introduction of new theoretical accounts. In concert with the accelerating development and applications of microprocessors that “have demanded not less but more of the human monitor” and “those who believe that just one more chip needs to be invented to automate the human out of the system” (Wiener, 1987, p. 735), a volume of tools are also growing for observation and data collection in instrumented vehicles, field operational studies, and naturalistic driving (Dingus et al., 2006, Eenink et al., 2014, McGehee et al., 2007, Regan et al., 2012, Stutts et al., 2005, Tivesten and Dozza, 2014, Victor et al., 2010). Collectively, such studies could begin to provide exactly the wealth of real-world operational knowledge needed to bridge theory and practice (e.g., Wiener, 1987). Furthermore, they typify and support emerging theoretical perspectives, that is, situated cognition, that posit knowledge as inseparable from doing by being situated in activity bound to social, cultural, and physical contexts (Robbins & Aydede, 2008).

Interestingly, with accelerating advances in computation (Moore, 1965), telecommunication, and Internet connectivity technology, nothing should inherently prohibit such real-world data from entering into laboratories and like areas of greater control and manipulation. For example, augmented reality and other blended designs might be an appealing approach (Hancock & Sheridan, 2011, chap. 4) as well as widely publically available and diverse driving video data sets (e.g., YouTube DashCam videos). Overall, in parallel with a growing popularity of debunking myths of “good” and “bad” drivers (Arnstein & Arnstein, 2005), future driving vigilance research efforts might benefit from following lines of cognitive and work domain analyses well used by many other domains (Rasmussen et al., 1994, Vicente, 1999), along with critical re-consideration of fundamental driving attention and distraction paradigms (Hancock, 2013b, Kircher and Ahlstrom, 2015) and direct consideration of real-world conditions and constraints typically under-represented in simulator studies, including the allowance of terminating/modulating vigilance task performance at one’s own intrinsic will rather than external compulsion that fixes down attention otherwise left free to vary (Hancock, 2013a, Scerbo, 2001).

From naturalistic driving studies, evidence is only recently emerging that safety risks associated with cell-phone use are considerably smaller than previously believed (Fisher, Caird, Rizzo, & Lee, 2011, chap. 1) by distinguishing between talking/listening vs. reaching/dialing cell-phone aspects and by comparing relative to other higher risk factors like drowsiness and specific environmental situations like intersections and increased traffic densities (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). Future studies may even begin to address the possibility of cell-phone use as a benefit, for example, as voluntary countermeasure to reduced alertness (Victor et al., 2015). The constant maintenance of some prescribed and pre-determined level of driving vigilance may itself also be worth challenging or in the least worth re-visiting provided more specific detailing of the situational features included in actual driving activity. Indeed, the lack of consensus from the

present analysis of driver vigilance operationalization may be viewed as support for reversals or at least re-examinations regarding assumptions or requirements of how drivers should, and/or how they actually do perceive and respond while driving.

Acknowledgments

The research presented in this paper was conducted within the project HFAuto – Human Factors of Automated Driving (PITN-GA-2013-605817). The authors would like to thank Miltos Kyriakidis, Laura Manca, Zhenji Lu, Natalia Kovácsová, and Joel Gonçalves for their instrumental efforts and contributions for the present analysis and Dimitra Dodou for her aid in the preparation of the manuscript

References

Note: References marked with an asterisk indicate publications included in the present analysis

*Amato, J., Marie, S., Lelong-Boulouard, V., Paillet-Loilier, M., Berthelon, C., Coquerel, A., et al. (2013). Effects of three therapeutic doses of codeine/paracetamol on driving performance, a psychomotor vigilance test, and subjective feelings. *Psychopharmacology*, 228, pp. 309-320. doi:10.1007/s00213-013-3035-7

Arnstein, Z. & Arnstein, L. (2005). *The bad driver's handbook: Hundreds of simple maneuvers to frustrate, annoy, and endanger those around you*. Santa Monica, CA: Santa Monica Press LLC.

*Atchley, P. & Chan, P. (2011). Potential benefits and costs of concurrent task engagement to maintain vigilance: A driving simulator investigation. *Human Factors*, 53(1), pp. 3-12. doi:10.1177/0018720810391215

Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), pp. 775-779. doi:10.1016/0005-1098(83)90046-8

Blincoe, L., Miller, T., Zaloshnja, E., & Lawrence, B. (2015). *The economic and societal impact of motor vehicle crashes, 2010. (Revised)* (Report No DOT HS 812 013). Washington, DC: National Highway Traffic Safety Administration (NHTSA). doi:10.1016/j.jannemergmed.2015.06.011

Bloomer, R. (1962). Perceptual defense and vigilance and driving safety. *Traffic Quarterly*, 16(4), pp. 549-558.

*Boadle, J. (1976). Vigilance and simulated night driving. *Ergonomics*, 19(2), pp. 217-225. doi:10.1080/00140137608931533

*Bonnefond, A., Roge, J., & Muzet, A. (2006). Behavioral reactivation and subjective assessment of the state of vigilance: Application to simulated car driving. *International Journal of Occupational Safety and Ergonomics*, 12(3), pp. 221-229. doi:10.1080/10803548.2006.11076684

Boverie, S., Giralt, A., & Le Quellec, J. (2008). Diagnostic fusion for in vehicle driver vigilance assessment. *Proceedings of the 17th World Congress of The International Federation of Automatic Country*. Seoul, Korea, Jul. 6-11. doi:10.3182/20080706-5-kr-1001.02169

*Brice, C. & Smith, A. (2001). The effects of caffeine on simulated driving, subjective alertness and sustained attention. *Human Psychopharmacology*, 16(7), pp. 523-531. doi:10.1002/hup.327

Brown, I. (1994). Driver fatigue. *Human Factors*, 36(2), pp. 298-314.

*Brown, I., Simmonds, D., Tickner, A. (2007). Measurement of control skills, vigilance, and performance on a subsidiary task during 12 hours of car driving. *Ergonomics*, 10(6), pp. 665-673. doi:10.1080/00140136708930920

*Campagne, A., Pebayle, T., & Muzet, A. (2004). Correlation between driving errors and vigilance level: Influence of the driver's age. *Physiology & Behavior*, 80, pp. 515-524. doi:10.1016/j.phsbeh.2003.10.004

CARRS-Q (2013). State of the Road: Monotony & hypovigilance, a fact sheet of the Centre for Accident Research & Road Safety – Queensland.

*Chan, M. (2008). Benefits and cost of dual-tasking in a vigilance task: A laboratory and driving simulator investigation. (Master's thesis). University of Kansas, Lawrence, KS, USA.

*Chuang, C., Ko, L., Jung, T., & Lin, C. (2014). Kinesthesia in a sustained-attention driving task. *NeuroImage*, 91, pp. 187-202. doi:10.1016/j.neuroimage.2014.01.015

*Correa, A., Molina, E., & Sanabria, D. (2014). Effects of chronotype and time of day on the vigilance decrement during simulated driving. *Accident Analysis and Prevention*, 67, pp. 113-118. doi:10.1016/j.aap.2014.02.020

Craig, A. (1984). Human engineering: The control of vigilance. In J. Warm (Ed.), *Sustained attention in human performance*. Chichester, U.K.: Wiley, pp. 247 – 291.

*Dalton, B., Behm, D., & Kibebe, A. (2007). Effects of sound types and volumes on simulated driving, vigilance tasks and heart rate. *Occupational Ergonomics*, 7, pp. 153-168.

Davies, D., & Parasuraman, R. (1982). *The psychology of vigilance*. London, U.K.: Academic Press.

*Desai, A. & Haque, M. (2006). Vigilance monitoring for operator safety: A simulation study on highway driving. *Journal of Safety Research*, 37(2), pp. 139-147. doi:10.1016/j.jsr.2005.11.003

Dingus, T., Klauer, S., Neale, V., Petersen, A., Lee, S., Sudweeks, J., et al. (2006). The 100-Car naturalistic driving study, Phase II – Results of the 100-Car field experiment. *U.S. Department of Transportation, National Highway Traffic Safety Administration*, Technical Report DOT HS 810 593. doi:10.1037/e624282011-001

Eenink, R., Barnard, Y., Baumann, M., Augros, X., & Utesch, F. (2014). UDRIVE: the European naturalistic driving study. *Proceedings of the 2014 Transport Research Arena Conference*. Paris, France, Apr. 14-17.

Ervin, R., Bogard, S., & Fancher, P. (2000). Exploring implications of the deceleration authority of adaptive cruise control for driver vigilance. *Proceedings of the 7th World Congress on Intelligent Transport Systems*, Torino, Italy, Nov 6-9.

*Findley, L., Unverzagt, M., Guchu, R., Fabrizio, M., Buckner, J., & Suratt, P. (1995). Vigilance and automobile

accidents in patients with sleep apnea or narcolepsy. *CHEST Journal*, 108(3), pp. 619-624. doi:10.1378/chest.108.3.619

Fisher, D., Caird, J., Rizzo, M., Lee, J. (2011). Chapter 1, Handbook of driving Simulation for engineering, medicine, and psychology: An overview. In D. Fisher, M. Rizzo, J. Caird, & J. Lee (Eds.), *Handbook of driving simulation for engineering, medicine, and psychology*. Boca Raton, FL, USA: CRC Press, Taylor & Francis Group, LLC, pp. 1.1-1.16.

Frankmann, J. & Adams, J. (1962). Theories of vigilance. *Psychological Bulletin*, 59(4), pp. 257-272. doi:10.1037/h0046142

Gasser, T. & Westhoff, D. (2012). BASt-study: Definitions of automation and legal issues in Germany. TRB Road Vehicle Automation Workshop, Irvine, CA, USA.

*Giusti, A., Zocchi, C., & Rovetta, A. (2009). A noninvasive system for evaluating driver vigilance level examining both physiological and mechanical data. *IEEE Transactions on Intelligent Transportation Systems*, 10(1), pp. 127-134. doi:10.1109/its.2008.2011707

Gold, C., Dambock, D., Lorenz, L. & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society 57th Annual Meeting*, 57(1), pp. 1938-1942. doi:10.1177/1541931213571433

*Guillerman, R., Radziszewski, E., & Caille, J. (1978). Effects of carbon monoxide on performance in a vigilance task (automobile driving). In R. Thornton (Ed.). *Smoking Behavior: Physiological and Psychological Influences – Symposium Proceedings*. Churchill Livingstone.

Hancock, P., Mouloua, M., & Senders, J. (2008). On the philosophical foundations of the distracted driver and driving distraction. In M.A. Regan, J.D. Lee, & K.L. Young (Eds.). *Driver Distraction: Theory, Effects, and Mitigation*. CRC Press Taylor & Francis Group, Boca Raton, FL, USA, pp. 11-30.

Hancock, P., & Sheridan, T. (2011). Chapter 4, Handbook of driving Simulation for engineering, medicine, and psychology: An overview. In D. Fisher, M. Rizzo, J. Caird, & J. Lee (Eds.), *Handbook of driving simulation for engineering, medicine, and psychology*. Boca Raton, FL, USA: CRC Press, Taylor & Francis Group, LLC, pp. 4.1-4.11.

Hancock, P. (2013a). In search of vigilance: The problem of iatrogenically created psychological phenomena. *American Psychologist*, pp. 97-109. doi:10.1037/a0030214

Hancock, P. (2013b). Chapter 2, Driven to distraction and back again. In M. Regan, T. Victor, & J. Lee (Eds.). *Driver Distraction and Inattention: Advances in Research and Countermeasures*. Chichester, England: Ashgate, pp. 9-25.

Heikoop, D., de Winter, J., van Arem, B., & Stanton, N. (2015). Psychological constructs in highly automated driving: A consensus model and critical comment on construct proliferation. *Theoretical Issues in Ergonomics Science*. doi:10.1080/1463922X.2015.1101507

Holland, J. (1958). Human vigilance: The rate of observing an instrument is controlled by the schedule of signal detections. *Science*, 128(3315), pp. 61-67. doi:10.1126/science.128.3315.61

Horswill, M., & McKenna, F. (2004). Drivers' hazard perception ability: Situation awareness on the road. In S. Banbury, & S. Tremblay (Eds.). *A cognitive approach to situation awareness: Theory and Application*, Ashgate, Aldershot, UK, pp. 155-175.

*Howard, M., Jackson, M., Kennedy, G., Swann, P., Barnes, M. & Pierce, R. (2007). The interactive effects of extended wakefulness and low-dose alcohol on simulated driving and vigilance. *Sleep*, 30(10), pp. 1334-1340.

*Inkeri, H. (2010). Fatigue while driving in a car simulator: Effects on vigilance performance and autonomic skin conductance. (Master's thesis). University of Jyväskylä, Jyväskylä, Finland.

*Jayakumar, A., Novak, M., Faber, J., & Bouchner, P. (2014). Driver vigilance monitoring – Impact of long tunnels. In J. Breton, J. Quartieri, M. Guida, D. Guida, & C. Guarnaccia (Eds.), *Latest Trends in Energy, Environment, and Development – Proceedings of the 7th International Conference on Urban Planning and Transportation*. June 3-5, WSEAS Press, pp. 251 – 256.

Ji, Q., Zhu, Z., & Lan, P. (2004). Real-time noninvasive monitoring and prediction of driver fatigue. *IEEE Transactions on Vehicular Technology*, 53(4), pp. 1052-1068. doi:10.1109/tv.2004.830974

Kircher, K. & Ahlstrom, C. (in press). Minimum required attention – an operationalisation of driver distraction. *Proceedings of the 8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Salt Lake City, Utah, Jun 22-25, 2015.

Klauer, S., Dingus, T., Neale, V., Sudweeks, J., & Ramsey, D. (2006). The impact of driver inattention on near crash risk: An analysis using the 100-Car naturalistic driving study data. *U.S. Department of Transportation, National Highway Traffic Safety Administration*, Technical Report DOT HS 810 593. doi:10.1037/e729262011-001

*Lin, C., Chuang, C., Huang, C., Tsai, S., Lu, S., Chen, Y., et al. (2014). Wireless and wearable EEG system for

evaluating driver vigilance. *IEEE Transactions on Biomedical Circuits and Systems*, 8(2), pp. 165-176. doi:10.1109/tbcs.2014.2316224

*Lo, C. (2005). The impact of shift, circadian typology, and bright light exposure on sleepiness, vigilance, and driving performance in Hong Kong taxi drivers. (Master's thesis). The Chinese University of Hong Kong, China.

*Lucidi, F., Devoto, A., Bertini, M., Braibanti, P., & Violani, C. (2002). The effects of sleep debt on vigilance in young drivers: An education/research project in high schools. *Journal of Adolescence*, 25, pp. 405-414. doi:10.1006/jado.2002.0487

Mackie, R., & O'Hanlon, J. (1977). A study of the combined effects of extended driving and heat stress on driver arousal and performance. In R. R. Mackie (Ed.), *Vigilance: Theory, operational performance, and physiological correlates*. New York, New York, U.S.A.: Plenum, pp. 537-558. doi:10.1007/978-1-4684-2529-1_25

Mackie, R. (1984). Research relevance and the information glut. In F. Muckler (Ed.), *Human factors review 1984*. Santa Monica, CA, USA: Human Factors Society, pp. 1-11.

Mackworth, N. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1:1, pp. 6-21. doi:10.1080/17470214808416738

Mackworth, N. (1950). Researches on the measurement of human performance. *Medical Research Council Special Report Series 268*. London: HMSO. doi:10.1097/00007611-195111000-00022

McGehee, D., Carney, C., Raby, M., Lee, J., & Reyes, M. (2007). The impact of an event-triggered video intervention on rural teenage driving. *Proceedings of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. Stevenson, Washington, USA, Jul. 9-12.

*Mets, M., Volkerts, E., Dunnebier, E., de Senerpont Domis, L., Olivier, B., & Verster, J. (2008). Effects of seasonal allergic rhinitis on driving ability, memory functioning, sustained attention, and quality of life. *The Open Allergy Journal*, 1, pp.19-25. doi:10.2174/1874838400801010019

*Michael, R. & Meuter, R. (2006). Sustained attention and hypovigilance: The effect of environmental monotony on continuous task performance and implications for road safety. *Proceedings of the Australasian Road Safety Research, Policing, and Education Conference*, vol. 10, Monash University.

Moore, G. (1965). Cramming more components onto integrated circuits. *Electronics*, 38(6). doi:10.1109/n-ssc.2006.4785860

NHTSA (2008). National motor vehicle crash causation survey, Report to Congress. (Report No. DOT HS 811 059). Washington, DC: National Highway Traffic Safety Administration (NHTSA).

NHTSA (2013). Preliminary statement of policy concerning automated vehicles. United States National Highway Traffic Safety Administration (NHTSA). http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf

*O'Hanlon, J., Robbe, H., Vermeeren, A., van Leeuwen, C., & Danjou, P. (1998). Venlafaxine's effects on healthy volunteers' driving, psychomotor, and vigilance performance during 15-day fixed and incremental dosing regimens. *Journal of Clinical Psychopharmacology*, 18(3), pp. 212-221. doi:10.1097/00004714-199806000-00006

Parasuraman, R. (1979). Memory load and event rate control sensitivity decrements in sustained attention. *Science*, 205(4409), pp. 924-927. doi:10.1126/science.472714

*Pei, Y. & Lianzhen, W. (2013). Effect of driving duration and work schedules on vigilance level and driving performance of bus drivers. *Applied Mechanics and Materials*, 321-324, pp. 1518-1521. doi:10.4028/www.scientific.net/amm.321-324.1518

Pettitt, M., Burnett, G., & Stevens, A. (2005). Defining driver distraction. *Proceedings of the World Congress on Intelligent Transport Systems*, San Francisco, California, USA.

*Preece, M., Geffen, G. & Horswill, M. (2008). Are individuals recovering from mild traumatic brain injury vigilant drivers? Poster from the School of Psychology, University of Queensland, St Lucia, Australia.

Rasmussen, J., Pejtersen, L., & Goodstein, L. (1994). *Cognitive systems engineering*. Wiley International Press.

Regan, M., Hallett, C., & Gordon, C. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis and Prevention*, 43(5), pp. 1771-1781.

Regan, M., Williamson, A., Grzebieta, R., & Tao, L. (2012). Naturalistic driving studies: Literature review and planning for the Australian naturalistic driving study. In *Australasian College of Road Safety Conference 2012*, Sydney, New South Wales, Australia.

Robbins, P. & Aydede, M. (2008). A short primer on situated cognition. In P. Robbins & M. Aydede (Eds). *The Cambridge Handbook of Situated Cognition*. New York, New York, USA: Cambridge University Press, pp. 3-10. doi:10.1017/cbo9780511816826.001

*Roge, J., Pebayle, T., & Muzet, A. (2001). Variations of the level of vigilance and of behavioral activities during simulated automobile driving. *Accident Analysis and Prevention*, 33, pp. 181-186. doi:10.1016/s0001-4575(00)00029-4

*Roge, J., Pebayle, T., Kiehn, L., & Muzet, A. (2002). Alteration of the useful visual field as a function of state

of vigilance in simulated car driving. *Transportation Research Part F*, 5, pp.189-200. doi: 10.1016/s1369-8478(02)00017-7

Rosenbloom, T. & Wolf, Y. (2001). Signal detection in conditions of everyday life traffic dilemmas. *Accident Analysis and Prevention*, 34, pp. 763-772. doi:10.1016/s0001-4575(01)00076-8

*Ruiz, J. (2013). Measuring the three attentional networks in a vigilance context and their relationship with driving behaviour. (Doctoral thesis). University of Granada, Granada, Spain.

SAE (2014). Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. On-Road Automated Vehicle Standards Committee, SAE International. http://standards.sae.org/13016_201401

*Santana, A., Jammes, B., Boverie, S., Giralt, A., Poulard, H., & Thomas, J. (2003). Driver vigilance monitoring: New developments within the AWAKE project. Proceedings of the IFAC Intelligent Components and Instruments for Control Application conference. Aveiro, Portugal.

Sarter, M., Givens, B., & Bruno, J. (2001). The cognitive neuroscience of sustained attention: Where top down meets bottom-up. *Brain Research Reviews*, 35, pp. 146-160. doi:10.1016/s0165-0173(01)00044-3

Scerbo, M. (2001). Stress, workload, and boredom in vigilance: A problem and an answer. In P. Hancock & P. Desmond (Eds) *Stress, workload and fatigue – Human Factors in Transportation*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc. pp. 267-278.

*Schmidt, E., Kincses, W., Schrauf, M., Haufe, S., Schubert, R. & Curio, G. (2007). Assessing drivers' vigilance state during monotonous driving. *Proceedings of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. Stevenson, Washington, USA, Jul. 9-12.

*Schmidt, E., Schrauf, M., Simon, M., Fritzche, M., Buchner, A. & Kincses, W. (2009). Drivers' misjudgment of vigilance state during prolonged monotonous daytime driving. *Accident Analysis and Prevention*, 41, pp. 1087-1093. doi:10.1016/j.aap.2009.06.007

*Schmidt, E., Schrauf, M., Simon, M., Buchner, A., & Kincses, W. (2011). The short-term effect of verbally assessing drivers' state on vigilance indices during monotonous daytime driving. *Transportation Research Part F*, 14, pp. 251-260. doi:10.1016/j.trf.2011.01.005

Stanton, N. & Young, M. (2000). A proposed psychological model of driving automation. *Theoretical Issues in Ergonomics Science*, 1(4), pp. 315-331. doi:10.1080/14639220052399131

Steff, F., & Spradlin, H. (2000). Driver distraction, aggression and fatigue: A synthesis of the literature and guidelines for Michigan planning (Report UMTRI-2000-10). Ann Arbor, Michigan, USA. The University of Michigan Transport Research Institute.

Stroh, C. (1971). *Vigilance: The problem of sustained attention – International Series of Monographs in Experimental Psychology*. Oxford, England: Pergamon Press Ltd.

Stutts, J., Feagans, J., Reinfurt, D., Rodgman, E., Hamlett, C., Gish, K., et al. (2005). Drive's exposure to distractions in their natural driving environment. *Accident Analysis and Prevention*, 37, pp. 1093-1101. doi:10.1016/j.aap.2005.06.007

*Thiffault, P., & Bergeron, J. (2003). Monotony of road environment and driver fatigue: A simulator study. *Accident Analysis and Prevention*, 35, pp. 381-391. doi:10.1016/s0001-4575(02)00014-3

*Tippin, J., Sparks, J., & Rizzo, M. (2009). Visual vigilance in drivers with obstructive sleep apnea syndrome. *Journal of Psychosomatic Research*, 67, pp. 143-151. doi:10.1016/j.jpsychores.2009.03.015

Tivesten, E., & Dozza, M. (2014). Driving context and visual-manual phone tasks influence glance behaviour in naturalistic driving. *Transportation Research Part F*, 26, pp. 258-272. doi: 10.1016/j.trf.2014.08.004

*Ueno, A., Tei, S., Nonomura, T., & Inoue, Y. (2009). An analysis of saccadic eye movements and facial images for assessing vigilance levels during simulated driving. In D. Harris (Ed.). *Engineering Psychology and Cognitive Ergonomics*. Berlin, Germany: Springer-Verlag. pp. 451-460. doi:10.1007/978-3-642-02728-4_48

Vicente, K. (1999). *Cognitive work analysis: Towards safe, productive, and healthy computer-based work*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, Taylor & Francis.

Victor, T., Bargman, J., Hjalmdahl, M., Kircher, K., Svanberg, E., Hurtig, S., et al. (2010). Sweden-Michigan Naturalistic Field Operational Test (SeMiFOT) Phase 1: Final report. Vehicle and Traffic Safety Centre at Chalmers (SAFER). Report 2010:02, Project C3 SeMiFOT.

Victor, T., Dozza, M., Bargman, J., Boda, C., Engstrom, J., Flannagan, et al. (2015). Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk. Strategic Highway Research Program (SHRP 2 – Chap. 9, Conclusions and Recommendations), TRB Report S2-S08A-RW-1.

*Vrignon, J. & Rakotonirainy, A. (2007). Impact of subjective factors on driver vigilance: A driving simulator study. In *International Conference on Driver Behaviour and Training*, 3rd, 2007, Dublin, Ireland.

Warm, J., Dember, W., & Hancock, P. (1996). Chapter 9, Vigilance and workload in automated systems. In R.

Parasuraman & M. Mouloua (Eds). *Automation and human performance: Theory and applications. Human factors in transportation*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc. pp. 183-200.

Warm, J. & Parasuraman, R. (1987). Special issue preface, Vigilance: Basic and applied research. In honor of Dr. Michel Loeb. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 29(6), pp. 623-624.

Wetton, M., Horswill, M., Hatherly, C., Wood, J., Pachana, N., & Anstey, K. (2010). The development and validation of two complementary measures of drivers' hazard perception ability. *Accident Analysis & Prevention*, 42(4), pp. 1232-1239. doi:10.1016/j.aap.2010.01.017

Wiener, E. (1987). Application of vigilance research: Rare, medium or well done? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 29(6), pp. 725-736.

*Wyon, D., Wyon, I., & Norin, F. (1995). The effects of negative ionization on subjective symptom intensity and driver vigilance in a moving vehicle. *Indoor Air*, 5, pp. 179-188. doi:10.1111/j.1600-0668.1995.t01-1-00004.x

*Wyon, D., Wyon, I., & Norin, F. (1996). Effects of moderate heat stress on driver vigilance in a moving vehicle. *Ergonomics*, 39(1), pp. 61-75. doi:10.1080/00140139608964434

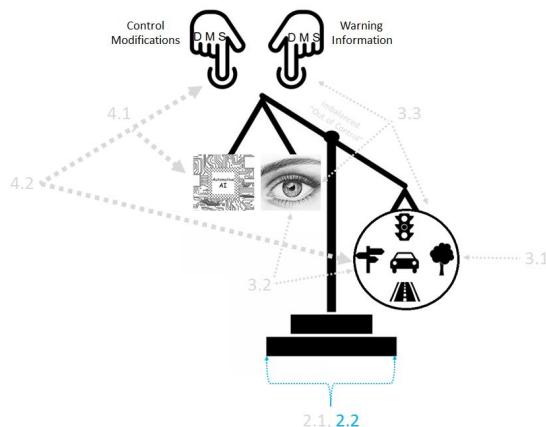
Young, K. & Regan, M. (2007). Driver distraction: A review of the literature. In I.J. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter & J.D. Irwin (Eds.). *Distracted driving*. Sydney, NSW: Australasian College of Road Safety. Pp. 379-405.

Appendix A. Supplementary Material

Supplementary data associated with this article can be found, in the online version, at:

- <http://dx.doi.org/10.1016/j.trf.2016.04.001>

Chap. 2.2 How to Keep Drivers Engaged while Supervising Driving Automation? A literature survey and categorization of six solution areas



In regards to the overall thesis big picture, this literature survey serves as a foundation for organizing previously proposed solutions to the problem of keeping the engagement of supervisors of automation (i.e., in general such that their lessons learned might be applied to the automated driving domain). The survey work generated six solution area themes with which independent raters exhibited better than chance agreement when tasked to apply the themes to categorize the conclusions found in 34 publications. The first three themes describe avoidance either in a hard sense or different versions of a soft stance: objective or subjective reductions in the supervisory control task. The latter three themes describe solutions under familiar learning theory paradigms in chronological order: behaviourism, cognitivism, and ecological constructivism. Cognitive followed by ecological themed solutions appear to be the most commonly proposed.

Adapted from:

Cabral, C.D.D., Eriksson, A., Dreger, F., Happee, R., & de Winter, J.C.F. (2019). How to keep drivers engaged while supervising driving automation? A literature survey and categorization of size solution areas. *Theoretical Issues in Ergonomics Science*, vol. 20(3), pgs. 332-365.

Abstract

This work aimed to organize recommendations for keeping people engaged during human supervision of driving automation, encouraging a safe and acceptable introduction of automated driving systems. First, heuristic knowledge of human factors, ergonomics, and psychological theory was used to propose solution areas to human supervisory control problems of sustained attention. Driving and non-driving research examples were drawn to substantiate the solution areas. Automotive manufactures might (1) avoid this supervisory role altogether, (2) reduce it in objective ways or (3) alter its subjective experiences, (4) utilize conditioning learning principles such as with gamification and/or selection/training techniques, (5) support internal driver cognitive processes and mental models and/or (6) leverage externally situated information regarding relations between the driver, the driving task, and the driving environment. Second, a cross-domain literature survey of influential human-automation interaction research was conducted for how to keep engagement/attention in supervisory control. The solution areas (via numeric theme codes) were found to be reliably applied from independent rater categorizations of research recommendations. Areas (5) and (6) were addressed by around 70% or more of the studies, areas (2) and (4) in around 50% of the studies, and areas (3) and (1) in less than around 20% and 5% respectively. The present contribution offers a guiding organizational framework towards improving human attention while supervising driving automation.

1. Background

1.1. Addressing human driving errors with automation technology

Traffic safety literature has predominately implicated human behaviour and cognition as principal factors that cause motor vehicle crashes and fatalities. Treat et al. (1979) performed 2,258 on-site and 420 in-depth accident investigations and found that human errors and deficiencies were a cause in at least 64% of accidents, and were a probable cause in about 90-93% of the investigated accidents. Treat et al. (1979) identified major human causes as including aspects such as improper lookout, excessive speed, inattention, improper evasive action, and internal distraction. The National Highway Traffic Safety Administration (NHTSA, 2008) conducted a nationwide survey of 5,471 crashes involving light passenger vehicles across a three year period (January 2005 to December 2007). NHTSA (2008) determined the critical reason for pre-crash events to be attributable to human drivers for 93% of the cases. Critical reasons attributed to the driver by NHTSA (2008) included recognition errors (inattention, internal and external distractions, inadequate surveillance, etc.), decision errors (driving aggressively, driving too fast, etc.), and performance errors (overcompensation, improper directional control, etc.).

Consequently, Advanced Driving Assistance Systems (ADAS) and Automated Driving Systems (ADS) are commonly motivated as solutions to address transportation safety problems of human errors (Kyriakidis et al., 2015; Gao et al., 2014; NHTSA, 2017). The Society of Automotive Engineers International (SAE) originally released a standard J3016_201401 (SAE, 2014) that conveyed an evolutionary staged approach of five successive levels of driving automation ranging from '*no automation*' to '*full automation*' (herein referred to as SAE Level 0-5). While the SAE standard has been revised several times to its most current version available as of June 2018 (SAE, 2018), its principal levels have been retained and continue to be a common reference point for the automotive automated/autonomous vehicles (AVs) research domain. Automotive manufacturers have already begun to release various SAE Level 2 '*Partial Automation*' systems within their on-market vehicles, which allow combined automatic execution of both lateral and longitudinal vehicle control under specific operational design domains. At SAE Level 2, drivers are still expected to complete object and event detection and response duties while retaining full responsibility as a fall-back to the driving automation (SAE, 2018).

1.2. New roles, new errors: Supervisors of mid-level driving automation

A complicating issue along the path to fully autonomous self-driving cars exists for the SAE Level 2 partial automation systems in regards to a state of driver supervisory engagement and retention of responsibility. Owners' manuals, manufacturer websites, and press releases of recent on-market SAE Level 2 systems were collected as background material to understand how the industry is presently addressing this issue. A sample of recently released SAE Level 2 driving automation system terminology and Human Machine Interfaces (HMI) regarding human disengagement is organized in Table 2.2.1. Notably, such concerns appear mostly in arguably passive (e.g., instructional guidelines and warnings), indirect (e.g., surrogate sensing of attention/involvement), and/or reactive (e.g., post-incident alerting) manners.

Most manufacturers kept their descriptions of driver engagement responsibilities and requirements during use of their SAE Level 2 systems at a higher level than commonly found in research communities (e.g., specifications of aberrant driver state terminology such as drowsiness,

distraction, inebriation). Instead, manufacturer examples included abstracted aspects like always being aware of and acting appropriately in traffic situations or being '*in control*'. Some notable specifics for the remaining driver responsibility include Mercedes' detailing of vehicle speed, braking, and staying in the lane (Mercedes-Benz, 2017, p. 177), a few statements from BMW that hands must be kept on the steering wheel (BMW, 2017), and repetitive remarks from Tesla regarding their hands-on requirements (Tesla, 2017, p. 73), including an entire sub-section entitled '*'Hold Steering Wheel'*' (Tesla, 2017, p. 74).

Table 2.2.1. Partially automated driving releases (~2017)[#]

Make	Model	System	Terms for driver state of engagement	Engagement Input ^a modality	Engagement Output ^b modality	Inattention escalation intervals
Volvo Cars	XC90 S90, V90	Pilot Assist II	attention, judgment	VLa VLn VMsc	AU VI TOC	0
GM, Cadillac	CT6	Driver Attention System (Super Cruise)	attention, awareness, supervision, engagement	VI	AU VI TA TOC	>1
Tesla	Model S Model X	Autopilot Tech Package v. 8.0	alert, safely, in control, hands-on, mindful, determine appropriate, be prepared	VLa	AU VI TOC	5
Audi	A4, Q7	Traffic Jam Assist	be in control, ready, responsible, assessing, attention	VLa VMsc	AU VI TOC	>1
BMW	750i 7 series	Active Driving Assistant Plus	be in control, responsible, correctly assess traffic situation, adjust the driving style to the traffic conditions, watch traffic closely, actively intervene, attentively	VLa	AU VI (TA) TOC	1
Infiniti	Q50S	Active Lane Control	be alert, drive safely, keep vehicle in traveling lane, control of vehicle, correct the vehicle's direction	(VLa)	(AU) (VI)	-1
Daimler, Mercedes- Benz	S65 AMG	Distronic Plus with Steering and Active Lane-Keeping Assist	adapt, aware, ensure, control, careful observation, be ready, maintain safety	VLa VMsc	AU VI (TA) TOC	1

^a Input modalities (vehicle from driver):

- VLa = vehicle lateral, steering, etc.
- VLn = vehicle longitudinal, brake, gas, etc.
- VMsc = vehicle misc., seat buckle, wait, door lock, etc.

^b Output modalities (vehicle to driver):

- AU = audio
- TA = tactile/haptic/vestibular
- VI = visual
- TOC = transition of control, change in functionality/level, etc.

sources of information

- Volvo Cars
 - http://volvornt.harte-hanks.com/manuals/2017/S90_OwnersManual_MY17_en-US_TP22301.pdf
 - http://volvo.custhelp.com/app/answers/detail/a_id/9769/~/new-features-available-as-of-november-2016
- GM, Cadillac
 - <http://media.gm.com/media/us/en/cadillac/news.detail.html/content/Pages/news/us/en/2017/apr/0410-supercruise.html>
 - https://www.youtube.com/watch?v=Shm3GY_JG-w
- Tesla
 - https://www.tesla.com/sites/default/files/model_s_owners_manual_north_america_en_us.pdf
- Audi
 - <http://ownersmanual.audiusa.com/>
 - http://www.audi.com/en/innovation/piloteddriving/assistance_systems.html
 - <https://www.youtube.com/watch?v=T8ESfICGnAc>
 - <https://www.youtube.com/watch?v=RMi4H4ybEkc>
- BMW
 - <https://www.bmwusa.com/owners-manuals.html>
 - <http://www.bmw.com/en/topics/fascination-bmw/connected-drive/developer-assistance.html>
 - <https://www.youtube.com/watch?v=RKAЕ-ANKIBY>
 - <https://www.youtube.com/watch?v=7fqXJscjzw>
- Infiniti
 - <https://owners.infinitiusa.com/content/manualsandguides/Q50/2017/2017-Q50-owner-manual-and-maintenance-info.pdf>
- Daimler, Mercedes-Benz
 - https://www.mbusa.com/mercedes/service_and_parts/owners_manuals#!year=2017&class=S-Sedan
 - http://techcenter.mercedes-benz.com/en/distronic_plus_steering_assist/detail.html
 - http://techcenter.mercedes-benz.com/en_ZA/steering-pilot/detail.html
- Unofficial demonstration/review reports
 - <https://www.youtube.com/watch?v=RjvI57BIDp0>
 - <https://www.caranddriver.com/features/semi-autonomous-cars-compared-tesla-vs-bmw-mercedes-and-infiniti-feature-2016-bmw-750i-xdrive-page-4>
 - https://www.youtube.com/watch?v=isZ3fSbE_pg
 - <https://www.youtube.com/watch?v=C7xV9rMajNo>

Across the various inputs that are interpreted as aberrant driver engagement/readiness (e.g., inadequate braking levels, unbuckled seatbelts, open doors, and driver facing cameras), the most common classification was that of measures associated with lateral vehicle control (i.e., steering wheel touch/torque and/or lane position). GM/Cadillac currently stands out as the only one so far to use a visual modality of a driver-facing camera to ascertain driver inattention. The consequential

output modalities of auditory, visual, and transitions of control (ToC) were found to be used by all manufacturers in their reactive HMI strategies. One manufacturer officially mentioned use of a tactile modality alert (GM/Cadillac) while a few others (Mercedes, BMW) were found in unofficial reports (MercBenzKing, 2016; Sherman, 2016).

By counting stages beyond a first warning (i.e., escalation intervals), Tesla was found to use the highest number of escalations in their reactive HMI. At least five escalations were observable from online Tesla owner videos (e.g., Black Tesla, 2016; Super Cars, 2017). Descriptions and approximated timings of the following escalations are in regards to coming after the initial warning of a grey filled textbox with wheel icon and '*Hold Steering Wheel*' message at the bottom of the dashboard instrument cluster.

- (1) +2 seconds after first warning - dashboard instrument cluster border pulses in white with an increasing rate
- (2) +15 seconds after first warning - one pair of two successive beeps
- (3) +25 seconds after first warning - two pairs of two successive beeps
- (4) +30 seconds after first warning - at the bottom of the instrument cluster, a red filled textbox plus triangle exclamation point icon with two line written messages of '*Autosteer Unavailable for the Rest of This Drive*' on line one, and '*Hold Steering Wheel to Drive Manually*' on line two in smaller font, along with a central image of two red forearm/hands holding a steering wheel that replaces the vehicle's lane positioning animation, the same previous pairs of successive beeps are repeated in a continuous manner; the vehicle gradually reduces speed
- (5) +37 seconds after first warning – all alerts from previous level remain, two yellow dots are added at the beginning of each forearm; the vehicle hazard blinkers are activated

A few manufacturers could be determined as having more than one escalation (GM/Cadillac, Audi), a few others as exactly one escalation (BMW, Daimler/Mercedes-Benz), and Volvo appeared to have a single first level/stage warning with no further escalation. Infiniti appeared to have no HMI reactive to driver disengagement/misuse of their Level 2 system (Active Lane Control). All but one manufacturer (Infiniti) were found to use at least the visual modality in their first stage of warning against driver disengagement.

2. Introduction of Solution Grouping Framework

2.1. Proactive solution strategies for human engagement in supervisory control

To complement the passive, indirect, and/or reactive approaches presently available in the aforementioned on-market industry examples, a set of proactive solution strategies towards human engagement in supervisory control might be helpful. Longstanding human factors and ergonomics principles have previously suggested risks in relying on humans as monitors of automated (e.g., invariant, predictable, monotonous, etc.) processes over extended periods (Greenlee et al., 2018; Hancock, 2017a; Molloy & Parasuraman, 1996; Bainbridge, 1983; Mackworth, 1950). Thus, it was

expected that many solutions might exist across the academic literature and could benefit from a qualitative framework for organizing trends and patterns in their recommendations.

A natural starting point to the difficulties in human supervisory control of driving automation is to avoid the supervisory role outright (e.g., skip SAE Level 2). Logically, softer versions of such a hard stance might also be realizable in either objective or subjective ways. Objectively, the amount of time or envelope of automated functionality could be reduced. Subjectively, the supervisory experience of responsibility could be refashioned with altered perceptions of the human's role towards shared or even fully manual authority. Furthermore, extensive research conducted under multiple paradigms of psychological theory might suggest approaches out of different schools of thought. The behaviourism paradigm centres around conditioning learning theories and suggests associative stimuli and/or stimulus-response pairing principles to promote the desired behaviour and discourage that which is undesirable. The cognitivism paradigm focuses on internal information processes and advises ways to support limited mental resources, representations, and awareness. Lastly, ecological approaches emphasize inclusion of external considerations of the task and the environment surrounding the worker/learner towards enhanced relational performance from a broader systems-level view. In summary, a grouping framework of six proactive solution areas is proposed to help answer the question '*How do we keep people engaged while supervising (driving) automation?*' In each case, the solution areas are introduced first in a general manner of various automation domains, before exemplifying relevancy specifically for engagement in supervisory control of driving automation.

Solution Area (1): Avoid the role of sustained human supervision of automation

- Suspend/repeal/skip levels of automation requiring human oversight and backup
 - *'just don't do it'*

Solution Area (2): Reduce the supervising role along an objective dimension

- Change the amount of time or envelope of automated operations
 - *'don't do it as much'*

Solution Area (3): Reduce the supervising role along a subjective dimension

- Share responsibilities and/or alter the end user experience and impressions
 - *'do it without drivers having to know about it'*

Solution Area (4): Support the supervising role from the behaviourism paradigm

- Condition the desired target behaviours through training and selection
 - *'make or find drivers who do it better'*

Solution Area (5): Support the supervising role from the dyadic cognitivism paradigm

- Inform designs to support cognitive processes and mental models
 - *'focus on internal mental constructs'*

Solution Area (6): Support the supervising role from the triadic ecological paradigm

- Inform designs to leverage external environment contexts and task considerations
 - *'focus on external task/environment factors'*

2.1.1. Solution Area (1): Avoid the role of human supervision of automation

The most parsimonious proactive solution could be to avoid subjecting drivers to the unnatural requirement of monitoring automated processes. Decades of human factors and ergonomics research have echoed that this is not something humans do well. A resounding result from Norman Mackworth (1948) was that despite instruction and motivation to succeed in a sustained attention task (used as an analogy to the critical vigilance of WWII radar operators watching and waiting for enemy target blips on their monitor screens), human detection performance dropped in relation to time-on-task. Thousands of reports have since been published on the challenges of human vigilance, also known as '*sustained attention*' (Frankmann & Adams, 1962; Craig, 1984; Cabrall et al., 2016). Bainbridge (1983) observed the irony that human supervisory errors are expected when operators are left to supervise an automated process put in place to resolve manual control errors. Humans were described as deficient compared to machines in prolonged routine monitoring tasks, as seen in the MABA-MABA (Men Are Better At – Machines Are Better At) list by Fitts (1951), and such characterizations persist today (De Winter & Dodou, 2011). In a review of automation-related aircraft accidents, Wiener and Curry (1980) suggested that it is highly questionable to assume that system safety is always enhanced by allocating functions to automatic devices rather than human operators. They instead consider first-hand whether a function *should* be automated rather than simply proceeding because it *can* be.

Driver responses have been found to be negatively impacted when having to respond to simulated automation failures while supervising combined automatic lateral and longitudinal driving control (De Waard et al., 1999; Stanton et al., 2001; Strand et al., 2014). From elaborated operator sequence diagram models, Banks et al. (2014) indicated that far from reducing driver workload, additional sub-system tasks associated with monitoring driving automation actually would increase cognitive loads on a driver. Banks et al. (2018) analysed on-road video observations of participants operating a Tesla Model S in Autopilot mode (i.e., SAE Level 2 driving automation). Their analysis suggested that '*drivers are not being properly supported in adhering to their new monitoring responsibilities and instead demonstrate behaviour indicative of complacency and over-trust*'. Accordingly, Banks et al. (2018) discussed a possibility that certain levels of driving automation (DM, driver monitoring) need not be implemented even if they are feasible from a technical point of view, and that a simplified set of roles of only DD (driver driving) and DND (driver not driving) could be preferred from a human factors role/responsibility point of view.

'...it seems more appropriate at the time to accept that the DD and the DND) roles are the only two viable options that can fully protect the role of the human within automated driving systems. This in turn means that either the human driver should remain in control of longitudinal and/or lateral aspects of control (i.e., one of the other) or they are removed entirely from the control-feedback loop (essentially moving straight to SAE 4)'. (p. 144).

2.1.2. Solution Area (2): Reduce the role along an objective dimension

In the mid-1990s, several key studies suggested a less strict avoidance approach in the human supervision of automation. Various schemes for alternating periods of manual and automated control were investigated (Parasuraman et al., 1996; Scallen et al., 1995; Endsley & Kiris, 1995). In Parasuraman et al. (1996), adaptive control conditions where control was temporally returned to a human operator showed subsequent increases in monitoring performance compared to a non-adaptive full automated condition. In Scallen et al. (1995), adaptive switching between manual and automated control was investigated at short time scale intervals (i.e., 15, 30, and 60 seconds).

Objective performance data indicated better performance with shorter rather than longer cycles. However, such benefits were associated with increased workload during the shorter cycle durations (i.e., the participants did better only at the cost of working harder and prioritizing a specific sub task). Thus, the authors concluded that if the goal of the operator is to maintain consistency '*on all sub-tasks, at all times*' then the performance immediately following episodes of short automation warrants particular concern: i.e., '*the results support the contention that excessively short cycles of automation prove disruptive to performance in multi-task conditions*'. In Endsley and Kiris (1995) the level of automated control was investigated. Rather than manipulating the length of time of automated control, a shift from human active to passive processing was deemed responsible for decreased situation awareness and response time performance. Manual control response times immediately following an automation failure were observably slower compared to baseline manual control periods. However, the effect was less severe under partial automation conditions compared to the full automation condition.

In Merat et al. (2014), a motion-based driving simulator experiment study was conducted with adaptive automation. They compared a predictable fixed schedule for triggering ToC to manual control with a real-time criterion which switched to manual based on durations of drivers looking away from the forward roadway. The authors concluded that better vehicular control performance was achieved when the automated to manual ToC was '*predictable and based on a fixed time*'.

2.1.3. Solution Area (3): Reduce the role along a subjective dimension

Rather than altering the objective amount of automated aid as in solution area (2), automation system design can also focus on the driver's psychological subjective experience or perception of responsibility and/or capability. In other words, manual human operator behaviour is not replaced in solution area (3) but augmented, extended, and/or accommodated. Such subjective shaping might take the form either as help (e.g., automatic backup) or even as hindrance (e.g., to provoke positive adaptive responses). Schutte (1999) introduced the concept of '*complementation*' to describe technology that is designed to enhance humans by augmenting their innate manual control skills and abilities rather than to replace them. With such complementary technology, '*many of the tasks that could be automated (i.e., performed solely by technology) are deliberately not automated so that the human remains involved in the task. This involvement must be meaningful rather than simply "doing something" or "busy work"*' (Schutte, 1999, p. 116., emphasis added). Flemisch et al. (2016) relayed similar theoretical concepts and design approaches where both the human and the machine should act together at the same time under a '*plethora*' of names, such as shared control, cooperative control, human-machine cooperation, cooperative automation, collaborative control, co-active design, etc. Young & Stanton (2002) proposed a Malleable Attentional Resources Theory positing that the size of relevant attentional resource pools can temporally adapt to changes in task demands (within limits). Thus, cognitive resources may actually be able to shrink/grow to accommodate various decreases/increases in *perceived* demands (e.g., even while retaining objective protections in the background).

Janssen (2016) evaluated simulated automated driving as a backup and found improved lateral performance and user acceptance (workload and acceptance) compared to adaptive automated-to-manual ToC. Mulder et al. (2012) improved safety performance and decreased steering variation in a fixed-base driving simulator through the use of haptic shared control. By requiring and retaining some level of active control from the human driver (i.e., amplification of a suggested torque), the

shared control model was expected by Mulder et al. (2012) to maintain some levels of engagement, situation awareness, and skill as compared to the supervisory control of automation.

A concept of promoting increased care in driving from the end-user by a seemingly reductive or even counter-productive human automation interface design can be found in Norman (2007). In order to keep human drivers informed and attentive, the proposition suggested that more requirements for human participation might be presented than is really needed. In other words, an automated driving system can encourage more attention from the human supervisor by giving an appearance of being less capable, of doing less, or even doing the wrong thing. Norman (2007) exemplified this framework of '*reverse risk compensation*' by reference to Hans Monderman (1945-2008) and then to Elliot et al. (2003). In Monderman's designs, the demarcations, rules, and right of ways of a designed traffic system are purposefully diminished/removed in favour of shared spaces. The idea is to provoke end-users (drivers, pedestrians, cyclists, etc.) to collectively combat complacency and over-reliance on rules/assumptions by being forced to look out for themselves (and one another). Norman (2007) cited results from Elliot et al. (2003) where artificial increases in perceived uncertainty resulted in driver adoption of safer behaviours such as increased information seeking and heightened awareness. In sum, Norman (2007) described an interesting potential of designed automated processes in futuristic cars where there could be an approach of shaping psychological experiences.

'...we can control not only how a car behaves but also how it feels to the driver. As a result, we could do a better job of coupling the driver to the situation, in a natural manner, without requiring signals that need to be interpreted, deciphered, and acted upon ... The neat thing about smart technology is that we could provide precise, accurate control, even while giving the driver the perception of loose, wobbly controllability'. (p. 83).

2.1.4. Solution Area (4): Support the role from the behaviourism paradigm

A historical psychological perspective on shaping people to behave as desired can be traced back to the early 1900s behaviourism learning models of Ivan Petrovich Pavlov ('*classical conditioning*') and Burrhus Frederic Skinner ('*operant conditioning*'). Broadbent and Gregory (1965) attributed prolonged watch detriments to a shift in response criterion whereby operators might be better persuaded towards reacting to doubtful signals (e.g., manipulation of payoff). More recently, the term '*gamification*' has been defined as the '*use of game design elements in non-game contexts*' (Groh, 2012) and was recognized in positive and negative ways to exemplify conditional learning aspects (Terry, 2011). In gamification, interface designs utilize the mechanics and styles of games towards increased immersion. Related approaches include an emphasis on skills either acquired over practice (e.g., training focus) and/or from innate pre-dispositions (e.g., personnel selection, individual differences, etc.). Neuro-ergonomic approaches in Nelson et al. (2014) improved vigilance task performance via transcranial direct current stimulation. Parasuraman et al. (2014) identified a genotype associated with higher skill acquisition for executive function and supervisory control. Sarter and Woods (1993, p. 118) advised directions to support awareness through '*new approaches to training human supervisory controllers*', and Gopher (1991) suggested potential promise via the enhancement of '*skill at the control of attention*'.

Behaviouristic dispositions are also observable in the automotive domain concerning increased driver vigilance with ADAS. Similar to the aforementioned investigations of selection interest (e.g., neurological disposition for enhanced cognitive executive control), automotive research

recommendations have included the implementation of training programs and/or gamified concepts. This solution area aims to enhance operators without enough attentive skills, or executive control for sustained focus, to instead obtain such skill/focus via extra practice, immersion, and/or motivation. Diewald et al. (2013) reviewed '*gameful design*' and saw promise for its use for in-vehicle applications (e.g., navigation, safety, and fuel efficiency). For driving safety, virtual money/points and virtual avatar passengers were identified as rewards/punishments tied to onboard diagnostics of driving styles. In Lutteken et al. (2016), a simulated highly automated highway driving vehicle performed longitudinal and lateral control while the human driver controlled lane changes as a manager of consent. A gamified concept consisting of partner teaming, virtual currency points that could be earned/spent, and time scores was found to motivate and increase the desired cooperative driver behaviours. In a test-track study, Rudin-Brown and Parker (2004) found increased response times to a hazard detection task while using adaptive cruise control (ACC). Rudin-Brown and Parker (2004) concluded that response times to the ACC failure were related to drivers' locus of control and suggested driver awareness training as a potential preventive strategy that could minimize negative consequences with using novel ADAS. The TRAIN-ALL (European Commission co-funded) project had the objective to develop training schemes and scenarios for computer-based training in the use of new ADAS (Panou et al., 2010). Panou et al. (2010) evaluated various ADAS training simulations so that trainees would learn how to optimally use ADAS without overestimating their functionality and maintain appropriate knowledge of their limitations.

2.1.5. Solution Area (5): Support the role from the dyadic cognitivism paradigm

The internal human mind is the focus of solution area (5). The chapter '*The Human Information-Processor*' of Card et al. (1983) described a model of communication and information processing where '*Sensory information flows into Working Memory through the Perceptual Processor*', '*Working Memory consists of activated chunks in Long-Term Memory*', and '*The basic principle of operation*' consists of cycles of recognizing and acting (e.g., resulting in commands to a motor processor). In accord with this seminal work, cognitive user-centric interface design theory and practices (e.g., Johnson, 2010) have generally used metaphors and constructs to align content, structure, and functions of computerized systems with content, structure, and functions of human minds: attention (Sternberg, 1969; Posner, 1978), workload (Ogden et al., 1979, Moray, 1982), situation awareness (Endsley, 1995), (mental-spatial) proximity compatibility principle (Wickens & Carswell, 1995), and multiple (modality) resource theory (Wickens, 1980, 1984). Similar mentally focused accounts persist for the topic of sustained attention and monitoring. Parasuraman (1979) concluded that loads placed on attention and memory are what drive decrements in vigilance. See et al. (1995) argued for the addition of a sensory-cognitive distinction to the taxonomy of Parasuraman (1979), where it was emphasized that target stimuli that are (made to be) more cognitively familiar would reduce vigilance decrement consequences. Olson and Wuennenberg (1984) provided information recommendations for user interface design guidelines regarding supervisory control of Unmanned Aerial Vehicles (UAVs) in a list that covered cognitive topics of transparency, information access cost minimisation, projections, predictions, expectations, and end-user understanding of automation. Sheridan et al. (1986) described the importance of mental models in all functions of supervisory control, including aspects for monitoring (e.g., sources of state information, expected results of past actions, and likely causes of failures) and intervening (options and criteria for abort and for task completion). Lastly, the highly cited human trust of automation theory from Lee and See (2004) underscored arriving at *appropriate* trust via cognitive

aspects of users' mental models of automation: understandable algorithms, comprehensible intermediate results, purposes aligned to user goals, expectancies of reliability, and user intentions.

The importance of mental process components is shared by SAE Level 2 simulator studies (De Waard et al., 1999; Strand et al., 2014; Beggiato et al., 2015) and theoretical accounts (Beggiato et al., 2015; Li et al., 2012). De Waard et al. (1999) were concerned with reduced driver alertness and attention in the monotonous supervision of automated driving. They found emergency response complacency errors in about half of their participants, and advocated providing feedback warnings pertaining to automation failures (e.g., clear and salient status indicators). Strand et al. (2014) appealed to an account of situation awareness to explain their findings of higher levels of non-response as well as decreased minimum times to collision when simulated driving automation was increased from an ACC to an ACC plus automatic steering system. Beggiato et al. (2015) used both a driving simulator study (post-trial questionnaires and interviews as well as eye gaze behaviour) and an expert focus group to investigate information needs between SAE Levels 0, 2, and 3, where they found the second level to be more exhausting than the other conditions due to the continuous supervision task. Beggiato et al. (2015) concluded that in contrast to manual driving where needs are more oriented around driving-task related information, for partially and highly automated driving requested information is primarily focused on status, transparency, and comprehensibility of the automated system. Li et al. (2012) conducted a survey of recent works on cognitive cars and proposed a staged/levelled alignment of automation functions (e.g., perception enhancement, action suggestion, and function delegation) with driver-oriented processes (stimuli sensation, decision making, and action execution) (cf. Parasuraman et al., 2000; Eriksson et al., in press).

2.1.6. Solution Area (6): Support the role from the triadic ecological paradigm

A broad ecological systems view is represented by solution area (6). This perspective relates vigilance problems to an artificial separation of naturally coupled observation-action-environment ecologies. As an extension to information processing approaches, the chapter '*A Meaning Processing Approach*' of Bennett and Flach (2011) described a semiotics model dating back to work of Charles Peirce (1839-1914) that widens a dyadic human-computer paradigm into a triadic paradigm of human-computer-ecology with functionally adaptive rather than symbolically interpretive behaviour. Flach (2018) observed that minds tend to be situated, in the sense that they adapt to the constraints of situations (like the shape of water within a glass). Gibson (1979) promoted a theory of affordances not as properties of objects but as direct perception of ecological relations and constraints. Particularly in the chapter '*Locomotion and Manipulation*', Gibson (1979) suggested that the dichotomy of the "*mental*" apart from the "*physical*" is an ineffective fallacy. Gibson promotes units of direct perception to be not of things, but of actions with things. Moreover he conveys that such affordances are not available equally in some universal manner, but instead are relatively bounded in a holistic manner. Wickens and Kessel (1979) accounted for a manual control superiority because of a task ecology of continual sensing and correcting of errors together (active adaptation) where additional information (i.e., physical forces) is provided beyond those available from prolonged sensing alone without continual action. Neisser (1978) dismissed accounts of humans as passive serial information processors and instead promoted an indivisible and cyclic account of simultaneous processes. Thus, from such a point of view, vigilance tasks could be considered as problematic because of artificial assumptions and attempts to separate perception and action (i.e., thinking before acting, perceiving without acting, etc.) and to unnaturally isolate a state of knowledge at a singular specific point in time or sensory modality.

Such ecological approaches that emphasize the importance of direct perception and informed considerations of adaptation to specific work domains (tasks and situations) are evident in common across multiple human factors and psychological theories: cognitive systems engineering (Rasmussen et al., 1994), situation awareness design (Endsley et al., 2003), ecological psychology (Vicente and Rasmussen, 1990), situated cognition (Suchman, 1987), embodied minds (Gallagher, 2005), the embedded thesis (Brooks, 1991; O'Regan, 1992), and the extension thesis (Clark & Chalmers, 1998; Wilson, 2004). Flach (1990) promoted the importance of ecological considerations by emphasizing that humans naturally explore environments, and thus models of human control behaviour have been limited by the (frequently impoverished) environments under which they were developed. He relayed that an overly simple laboratory tracking task '*turns humans into a trivial machine*' and that real natural task environments (of motion, parallax, and optic arrays, etc.) are comparatively information rich with relevant '*invariants, constraints, or structure*'. Chiappe et al. (2015) supported a situated approach by observing that '*operators rely on interactions between internal and external representations to maintain their understanding of situations*' in contrast to traditional models that claim '*only if information is stored internally does it count as SA*'. Mosier et al. (2013) provided examples that the presence of traffic may affect the extent to which pilots interact with automation and the level of automation they choose and operational features such as time pressure, weather, and terrain may also change pilots' automation strategies as well as individual variables such as experience or fatigue. They found that vignette descriptions of different situational configurations of automation (clumsy vs. efficient), operator characteristics (professional vs. novice), and task constraints (time pressure, task disruptions) led pilots to different predictions of other pilots' behaviours and ratings of cognitive demands. Hutchins et al. (2013) promoted an integrated software system for capturing context through visualization and analysis of multiple streams of time-coded data, high-definition video, transcripts, paper notes, and eye gaze data in order to break through an '*analysis bottleneck*' regarding situated flight crew automation interaction activity. In an UAV vigilance and threat detection task, Gunn et al. (2005) recommended sensory formats and advanced cuing interfaces and accounted for the reduced workload levels they obtained via a pairing of detections to immediately meaningful consequential actions in a simulated real-world setting (i.e., shooting down a target in a military flight simulation) rather than responses devoid of meaning.

Leveraging external contextual information can be found in several recent driving automation theory and experimental studies. Lee and Seppelt (2009) convey that feedback alone is not sufficient for understanding without proper context, abstraction, and integration. Although technically an SAE Level 1 system, ACC also contains supervisory control aspects (i.e., monitoring of automated longitudinal control), and Stanton & Young (2005) concluded that ACC automation designs should depart from conventions that report only their own status, by offering predictive information that identifies cues in the world and relations of vehicle trajectories. Likewise, Seppelt and Lee (2007) promote and found benefits of an ecological interface design that makes limits and behaviour of ACC visible via emergent displays of continuous information (time headway, time to collision, and range rate) that relates the present vehicle to other vehicles across different dynamically evolving traffic contexts. In terms of an SAE Level 2 simulation, participants in Price et al. (2016) observed automated lateral and longitudinal control where vehicle capability was indicated via physically embodied lateral control algorithms (tighter/looser lane centre adherence) as opposed to via typical visual and auditory warnings. Consequentially, drivers' trust was found to be sensitive to such a situated communication of automation capability. Pijnenburg (2017) improved vigilance and decreased mental demand in simulated supervisory control of SAE Level 2

driving automation via a naturalistic interface that avoided arbitrary and static icon properties in its visual design. A recent theory of driving attention proposed not to assume distraction from the identification of specific activities alone but instead underscored a definition that requires relation in respects to a given situation (Kircher & Ahlstrom, 2017). After conducting several driver monitoring system (DMS) studies, a concluding recommendation from a work package deliverable of a human factors of automated driving consortium project was to '*incorporate situated/contextualized aspects into DSM systems*' (Cabral et al., 2017).

2.2. Literature Survey Aims

In the previous section, a qualitative grouping framework of six solution areas was introduced to identify trends and group proactive approaches towards human engagement while supervising automated processes. The aim of the following literature survey was to investigate whether the proposed solution areas might be represented in best practice recommendations and conclusions of influential and relevant works from a variety of human operator domains. Additionally, we aimed to identify trends between the solution areas: would some be more commonly found than others?; which might be more/less favoured by different domains?

3. Methods of Literature Survey

3.1. Inclusion Criteria

A scholarly research literature survey was conducted concerning the topic of keeping prolonged operator attention. In line with the terminology results of the automotive on-market survey (Table 2.2.1), our search terms were crafted to diminish potentially restrictive biases: of preferential terminology (vigilance, situation awareness, signal detection theory, trust, etc.), of operationalisation of performance (response/reaction time, fixations, etc.), of state (arousal, distraction, mental workload, etc.), or of specific techniques/applications (levels of automation, autonomous systems, adaptive automation, etc.). Instead, a more general Google Scholar search was performed with two presumably synonymous terms '*engagement*' and '*attention*'. The proactive term (i.e., '*keeping*') was included at the front of the queries to attempt to focus the literature survey away from reactive research/applications (e.g., concerning measurement paradigms).

(1) *keeping engagement in supervisory control*

(2) *keeping attention in supervisory control*

Google Scholar was used to reflect general access to semantically indexed returns from a broad set of resources as sorted for relevancy and influence in an automatic way. Literal search strings within more comprehensive coverage of specific repository resources were not presently pursued because the present survey was aimed initially for breadth and accessibility rather than database depth or prestige. Comparisons to a more traditional human-curated database (i.e., Web of Science) have concluded that Google Scholar has seen substantial expansion since its inception and that the majority of works indexed in Web of Science are available via Google Scholar (De Winter et al., 2014). Across various academic and industry research contexts, not all stakeholders might share equivalent repository reach, whereas Google Scholar is purposefully engendered as a disinterested and more even playing field. For such a democratic topic of driving safety risks while monitoring driving automation (i.e., that have already been released onto public roadways and might pose

dangers for everyone in general), organization of accessible guideline knowledge collectible from a broad-based Google Scholar resource seemed an appropriate first place methodological motivation ahead of future studies that might make use of more specific in-depth databases.

The 100 titles and abstracts of the first 50 results per each of the 2 search terms were reviewed to exclude work not pertaining to human-computer/automation research. Furthermore, several relevant and comprehensive review works that were returned in the search (e.g., Sheridan, 1992; Chen et al., 2011; Merat & Lee, 2012; etc.) were not included for categorization on the basis that their coverage was much wider than the present purposes of organizing succinct empirical recommendations. Exclusions were also made for works that appeared to focus more on promoting or explaining supervisory control levels or models of automation rather than concluding design strategies to the problem of operator vigilance while monitoring automated processes. One final text was excluded where raters had trouble applying a solution area on the basis that it dealt with remote human operation of a physical robotic manipulator. The research did not seem to share the same sense of human-automation supervisory control as seen in the other texts. The remaining set of 34 publications are listed in Appendix A by reverse chronological order.

3.2. Solution Area Categorizations via Numeric Theme Codes

To investigate the reliability of organizing the body of published literature with the proposed solution areas, confederate researchers (i.e., human factors PhD student (co-) authors on the present paper) were tasked as raters to independently categorize the conclusions of the retrieved research papers. For the sake of anonymity, the results of the three raters are reported with randomly generated pseudonym initials: AV, TX, and CO. Raters were provided an overview of the solution areas with numeric theme codes (i.e., Theme 1-6) and tasked with assigning a single top choice code for each of the publications of the inclusion set. The task was identified to the raters as *“to assign a provided theme code number to each of the provided publications texts based on what you perceive the best fit would be in regards to the authors’ conclusions (e.g., solution, strategy, guideline, recommendation)”*. Raters were also instructed to rank order any additional theme codes as needed. A survey rather than a deep reading was encouraged, where the raters were asked to sequentially bias their reading towards prioritized sections and continue via an additional as-needed basis (e.g., abstract, conclusions, discussion, results, methods, introduction, etc.) in order to determine the solution area that the author(s) could conceivably be most in favour of. A frequency weighting-scoring system per each theme code was devised where 1 point would be assigned for first choice responses, 0.5 points for second choice responses, and 0 points otherwise.

4. Results of Rater Categorizations

4.1. Inter-rater Reliability

First and second choice (where applicable) theme codes from each rater for each publication are presented in Appendix B. For first choice theme codes, statistical inter-rater Kappa agreement was computed via the online tool of Lowry (2018) with standard error computed in accordance with the simple estimate of Cohen (1960). The Kappa between AV and TX was 0.25, with a standard error of 0.11. The Kappa between AV and CO was 0.23, with a standard error of 0.11. The Kappa between TX and CO was 0.21, with a standard error of 0.09. Such Kappa statistic results (i.e., in the range of 0.21-0.40) may be interpreted as representing a ‘fair’ strength of agreement when benchmarked by the scale of Landis and Koch (1977) which qualitatively ranges across descriptors of ‘poor’, ‘slight’,

'fair', 'moderate', 'substantial', and 'almost perfect' for outcomes within six different possible quantitative ranges of Kappa values.

Initially suggestive of a low level of percentage agreement, only 6 out of the 34 publications received the same first choice coded theme categorization across all three raters. However, randomization functions were used to generate 3 chance response values (i.e., 1-6) for each of the 34 publications and repeated 100 different times. Thus, it was determined that the chance probability of achieving full way agreement for 6 or more publications was less than 1%. In comparison, random chance full agreement was observed for 0 publications to be 40%, for 1 publication to be 37%, for 2 publications to be 15%, for 3 publications to be 6%, for 4 publications to be 1%, for 5 publications to be 1%, and for 6 or more publications to be < 1%. Simulations with up to 1 million repetitions verified such a range of chance performance across 0 to 6 publications: 38%, 37%, 18%, 5%, 1%, < 1%, 0%.

Furthermore, matched categorizations between any 2 rather than all 3 of the raters was considered. As such, 27 out of the 34 publications received the same first choice coded theme categorization between at least 2 raters. As with the preceding full agreement analyses, random chance probabilities of two-way agreement were also computed from 100 sets of 3 random values for each of the 34 publications. The chance probability of achieving two-way categorization agreement for 27 or more publications was also determined to be less than 1%. In comparison, random chance two-way agreement was observed for between 31-34 publications to be less than 1%, for 26-30 publications to be less than 1%, for 21-25 publications to be 5%, for 16-20 publications to be 42%, for 11-15 publications to be 46%, for 6-10 publications to be 7% and for 5 or fewer publications to be less than 1%. Simulations with up to 50,000 repetitions verified such chance performance across the ranges of 31-34, 26-30, 21-25, 16-20, 11-15, 6-10, and 0-5 respectively as 0%, < 1%, 3%, 41%, 50%, 5%, and < 1%.

4.2. Theme Frequency

Weighted frequency scores (i.e., from aggregated first and second choice responses across raters) for each theme code and per each publication are listed in reverse chronological order in Table 2.2.2. Theme 5 appears to be the most common solution area, followed closely by 2 and 6. In contrast, Theme 1 appears to be the rarest, followed by Theme 3. While the majority of publications received heavy score weightings distributed across several themes, a highest likelihood single theme was recognizable for 28 of the 34 references (82%), as a result of the first and second choice rater aggregation scoring scheme. Theme 2 of objective reduction of amounts of human supervisory control of automation was found to be the most frequent first choice solution area labelled by 2 out of the 3 raters (i.e., AV and CO), whereas TX most often identified Theme 5 pertaining to support of internal cognitive processes and mental models. Theme 5 was also the most frequent second choice for TX and AV. Theme 6 regarding the use of external contexts and task considerations was the most frequent second choice of CO.

Table 2.2.2. Weighted frequency scores for aggregated first and second choices by each inter-rater for each publication reference. Lower/higher weights are lighter/heavier shaded. Highest weights per publication are outlined.

Ref ID	Weight of Theme 1	Weight of Theme 2	Weight of Theme 3	Weight of Theme 4	Weight of Theme 5	Weight of Theme 6
1	0.0	2.0	0.0	0.0	2.0	0.0
2	0.0	1.0	0.0	0.0	0.5	2.5
3	0.0	2.0	0.0	1.0	0.5	0.5
4	0.0	1.5	0.0	0.0	0.5	2.5
5	0.0	0.0	0.0	2.0	1.5	0.5
6	0.0	0.5	0.0	1.5	1.0	1.0
7	0.0	0.0	0.0	3.0	0.5	0.0
8	0.0	2.0	0.0	0.5	1.0	0.0
9	0.0	2.5	0.0	0.5	1.5	0.0
10	0.0	2.5	0.0	0.0	1.0	0.0
11	1.0	1.0	0.0	0.0	1.0	0.0
12	0.0	2.0	0.0	0.0	0.5	1.0
13	0.0	0.0	0.0	1.0	1.5	2.0
14	0.0	2.5	0.0	0.0	0.0	2.0
15	0.0	3.0	0.5	0.0	0.0	0.5
16	0.0	0.0	0.0	3.0	0.0	0.5
17	0.0	0.0	0.0	0.5	1.0	3.0
18	0.0	2.0	0.0	0.0	1.0	0.5
19	0.0	1.0	0.0	0.0	1.0	1.0
20	0.0	0.0	1.0	0.0	1.0	2.0
21	0.0	2.0	0.5	0.0	1.0	0.5
22	0.0	0.0	1.5	0.0	1.0	1.5
23	0.0	1.0	0.0	2.0	0.5	0.0
24	2.0	2.5	0.0	0.0	0.0	0.0
25	0.0	1.0	0.0	2.0	0.5	0.0
26	0.0	0.0	0.0	0.5	2.0	1.5
27	0.0	0.0	0.0	3.0	0.5	0.5
28	0.0	0.0	0.0	0.0	1.0	2.5
29	0.0	0.0	1.0	0.0	0.5	2.0
30	0.0	0.0	0.0	1.0	3.0	0.0
31	0.0	0.0	1.0	0.0	1.5	1.0
32	0.0	1.0	0.0	0.0	1.5	1.5
33	0.0	0.0	0.0	0.0	2.0	1.5
34	0.0	0.5	0.0	2.0	2.0	0.0
Total:	3.0	33.5	5.5	23.5	34.0	32.0

All publications of the included thematic analysis set were informally organized into primary operational domain(s) of concern (i.e., what job or service was the human supervisory control of automation investigated in). Most likely solution areas from weighted raters' first and second choice applied theme codes were determined per publication. Domains and most likely themes are combined in reverse chronological order in Table 2.2.3. In general, it can be observed that for the included publications, the domain areas have shifted over the decades from more general laboratory and basic research and power processing plants towards more mobile vehicle/missile applications and most recently especially with remotely operated vehicles. Although of limited sample size, some general domain trends might be observed. For example, it appears that uninhabited aerial vehicle (UAV) operations predominately favoured Theme 2 with also some consideration for Theme 6. In contrast, uninhabited ground vehicle (UGV) operations presently indicated only Theme 4. Earlier work with space, power plants, and general basic research showed a mix mostly of Themes 5 and 6. Aviation areas with pilots and air traffic control had a split of Themes 4 and 5. Missile air defence consisted of Theme 4 and Theme 2. Lastly, two automobile studies were present in the returned results: the first involving a fairly abstracted driving decision task (with a resulting likely categorization of Theme 2), and the second evidencing a split categorical rating assignment between Theme 2 and Theme 5.

Table 2.2.3. Primary operator domains of publications with identified likely thematic solution category from aggregate inter-rater first and second choice weighted scores. U(x)V = uninhabited vehicles, robots; UAV = uninhabited aerial vehicles; UGV = uninhabited ground vehicles; USV = uninhabited surface vehicles, ships; UUV = uninhabited underwater vehicles; Pilot = flight-deck, cockpit; ATC = ground-based air traffic control; Missile = air defense command and control; Automobile = automotive cars, trucks, etc.; Naval vessel = battleship, aircraft carrier, etc.; Space = spacecraft, satellites, etc.; Power plant = hydro, nuclear, electric, gas, oil, etc.; General = laboratory, basic research; Radar = military asset defence of airfield, ship, etc.; ComCon = general military command/control, tactical operations

Ref ID	U(x)V	UAV	UGV	USV	UUV	Pilot	ATC	Missile	Auto-mobile	Naval Vessel	Space	Power plant	General	Radar	ComCon
1												2/5			
2		6													
3	2	2				2									
4	6														
5								4							
6								4							
7									4						
8	2	2			2										
9	2	2				2									
10					2										
11		1/2/5													
12									2						
13		6													
14		2													
15		2										2			
16									4						
17		6													
18	2	2													
19	2/5/6														
20												6			
21	2	2													
22							3/6					3/6			
23												4			
24									2						
25						4									
26						5									
27								4							
28											6				
29							6			6					
30						5							6		
31											5	5			
32												5/6			
33										5					
34												4/5			

5. Discussion

5.1. Evolution of Cross Domain Concern

With a proliferation of automation also comes an increase in human supervision of automation (Sheridan, 1992) because automation does not simply replace but changes human activity. Such changes often evolve in ways unintended or unanticipated by automation designers and have been predominately regarded in a negative sense as in '*misuse*', '*disuse*', and '*abuse*' (Parasuraman & Riley, 1997) and/or as '*ironies*' (Bainbridge, 1983). Whether or not significant human supervisory problems will manifest in a proliferation commiserate with automation propagation is likely to be a function of the automation's reliability in the handling of the problems inherent in its' domain area. Human supervisors of automation are needed not only because a component might fail (e.g., electrical glitch) but also because the situation might exceed the automatic programming. Originally, computers and their programs were physically much larger and constrained to determinable locations within predictable and enclosed environments. As computers have become physically smaller their automated applications could be more practically incorporated into vehicles. Vehicles, however literally move across time and space and hence are subject to many environmental variants. Advances in supervisory control automation have been originally appropriate and suitable to vast expanse domains (outer space, the oceans, the sky) because they are difficult for humans to safely and commonly inhabit. Thus, such domains typically suffer from impoverished infrastructures and are subject to signal transmission latencies where automation must close some loops itself. Such automatic closures are benefited further by the absence of masses of people because compared to machines, people create a lot of noise and uncertainty with many different kinds of unpredictable and/or imprecise behaviours.

Likewise, driving automation was first showcased on highly structured freeways (Ellingwood, 1996), out in the desert and within a staged urban environment on a closed air force base (DARPA, 2014) before progressing towards more open operational design domains. Subsequently, driving automation market penetration has tended to begin first within more closed campus sites and scenarios with lower levels of uncertainty (e.g., interstate expressways) before proceeding into other contexts of increasing uncertainty and/or complexity (e.g., state highways, rural roads, and urban areas). Thus, while the present search terms for keeping attention/engagement in supervisory control returned only two studies in the automotive area, more might be expected in the future to the extent that 1) automated vehicles continue to need human supervisors (e.g., how structured and predictable vs. messy and uncertain are the areas in which they drive) and 2) how much attention/engagement of human supervisors of automated driving might be expected to wane or waver.

5.2. Convergence and Contribution

When restricted to a single choice, seemingly few applied theme codes were found to be in common agreement across all three independent raters. However, non-chance agreement was still obtained both in terms of standard inter-rater reliability Kappa statistics and percentage agreement analyses. Furthermore, thematic categorization agreement was enhanced by the allowance of rater second choices, which seems plausible, as empirical research conclusions can of course be of compounding nature. For example, Stanton et al. (2001) address the design of future ADAS by advocating for future research that '*could take any of the following forms: not to automate, not to automate until technology becomes more intelligent, to pursue dynamic allocation of function, to use technology to monitor and advise rather than replace, to use technology to assist and provide*

additional feedback rather than replace, to automate wherever possible'. Saffarian et al. (2012) proposed several design solution areas for automated driving: shared control, adaptive automation, improved information/feedback, and new training methods. Specifically for the topic of SAE Level 2 'partially automated driving', Casner et al. (2016) lament their expectations for vigilance problems in their conclusions that '*Today, we have accidents that result when drivers are caught unaware. Tomorrow, we will have accidents that result when drivers are caught even more unaware*'. Furthermore, they anticipate dramatic safety enhancements are possible when automated systems share the control loop (such as in backup systems like brake-assist and lane-keeping assistance) or adaptively take it as needed from degraded driver states (i.e., distraction, anger, intoxication). Casner et al. (2016) also conclude that designers of driver interfaces will not only have to make automated processes more transparent, simple, and clear, they might also periodically involve the driver with manual control to keep up their skills, wakefulness, and/or attentiveness. Lastly, Seppelt and Victor (2016) suggest new designs (better feedback and environment attention-orienting cues) as well as '*shared driving wherein the driver understands his/her role to be responsible and in control for driving*' and/or fully responsible driving automation that operates without any expectation that the human driver will serve as a fall-back.

The proposed solution areas overlap with many of the compounded review conclusions above from Stanton et al. (2001), Saffarian et al. (2012), Casner et al. (2016), and Seppelt and Victor (2016). From the present literature survey, what is added is a grouping framework that might more fully encapsulate the conclusions of empirical results from both the broad body of human factors, ergonomics, and learning theory as well as human driving automation interaction research. Furthermore, the solution areas were purposefully organized in a hopefully digestible and memorable way. The first three themes describe avoidance either in a hard sense or different versions of a soft stance: objective or subjective reductions. The latter three themes describe solutions under familiar learning theory paradigms in chronological order: behaviourism, cognitivism, and ecological constructivism.

Identifying a 'best' or 'preferred' theme of proactive strategy is not expected to be a discretely resolvable answer. Instead, the relative advantages and disadvantages should probably best be reflected upon in light of contextual considerations. Furthermore, due to their qualitative nature, the themes are not directly orthogonal from one another. Themes 2 and 3 could be conceived of as softer avoidance versions of a stricter skip-over stance of Theme 1. Theme 6 can be seen to expand from Theme 5 not as an opposing contrast but as an elevating extension that can still subsume cognitive and human-centred concepts. Themes 5, 2, and 6 were the top three most common solution areas found in the present survey.

5.2.1. Solution Area (1): Avoid the role of human supervision of automation

For Theme 1, it might be easier to hold close to a viewpoint of avoiding supervisory control of automation in theoretical or laboratory-oriented research. A sizeable body of human factors and ergonomics science literature supports such a standpoint that human bias and error is not necessarily removed via the introduction of automation, but instead, humans can generally be shown to be poor monitors of automation. However, industry examples also exist of both traditional and start-up automotive manufacturers (i.e., Ford and Waymo) opting to skip mid-level driving automation where a human is required to continuously supervise the processes (Ayre, 2017; Szymkowski, 2017). The low coverage of this theme in the present survey (see Table 2.2.2) is

probably more an artefact of the present survey rather than evidence of its unimportance or non-viability—more discussion is provided in a separate limitations section.

5.2.2. Solution Area (2): Reduce the role along an objective dimension

Regarding Theme 2, temporal restrictions based upon scheduled durations of automation use might be a practical starting place to initially implement mechanisms to reduce the objective amount of human supervision of driving automation. For combatting fatigue associated with conventional driving control during long trips, many modern day vehicles come equipped with timing safety features. Such rest reminders function by counting the elapsed time and/or distance of a single extended trip (e.g., hours of continuous operation since ignition on) and consequently warn/alert the driver for the sake of seeking a break or rest period. Because time on task has been traditionally identified as a major contributing factor to vigilance problems (Mackworth, 1948; Teichner, 1974; Greenlee et al., 2018), time-based break warnings and/or restrictions as with general driving fatigue countermeasures, might be practically worthwhile to apply on scales specific for human supervisory monitoring of SAE Level 2 driving automation. Compared to other contributing components to vigilance decrements (cf. Cabrall et al., 2016), the duration of watch period is expected to be an attractive dimension for human-automation interaction system designers due to its intuitive and simplistic operationalization even in spite of its potential to interact with other vigilance factors.

5.2.3. Solution Area (3): Reduce the role along a subjective dimension

Theme 3 of altering the perception towards increased danger or uncertainty and thus necessitating greater care from end-users could be problematic for automotive manufacturers that would reasonably expect to maintain positive rather than negative attributions of their products and services. However, an altered experience might carefully be crafted to direct attribution of uncertainty away from the vehicle and towards aspects of the environment or others (see Norman, 2007, pp. 83-84). For example, advanced driving automation of SAE Level 2 (simultaneous lateral and longitudinal control) might operate on an implicit level to support a driver who believes that he/she alone has control authority/responsibility (e.g., in line with how previous lower level driver assistance systems such as electronic stability control have been successfully deployed in the background). Discussion of its relatively low amount of coverage in the present survey (see Table 2.2.2) is provided in a separate limitations section.

5.2.4. Solution Area (4): Support the role from the behaviourism paradigm

Theme 4 is perhaps the most widely known in the general population and especially that behaviouristic aspect of manipulating or shaping behaviour through rewards and punishments. Caution, however, is warranted, as effects have been previously shown to be limited in lasting power and reach. For example, Parasuraman & Giambra (1991) found that while training and experience can help to reduce vigilance decrements, its benefits were not as observable in older populations: practice alone is insufficient to eliminate age differences. Notably, elderly populations are commonly regarded as primary users and beneficiaries of automated/autonomous ADAS (cf. Hawkins 2018). Furthermore, the practical viability of Theme 4 should be noted with consideration of the fact that a large proportion of the vigilance decrement phenomena exhibited in historic experiments was undertaken by young, highly trained, and motivated operators. By comparison, the present literature survey was concerned with uncovering proactive knowledge further generalizable and applicable to laypeople who might not be used to or amenable to rigours of

professional training when it comes to driving (e.g., recurrent training, reading of documentation, attention to help resource media/material, etc.).

5.2.5. Solution Area (5): Support the role from the dyadic cognitivism paradigm

Theme 5 cognitive science approaches have become prominent and favoured over the last few generations. Established human-automation research guideline approaches are on the rise (i.e., information processing models, awareness/attention, user/human centred design, etc.) alongside the popular success of companies like Google that promote their top maxim as '*Focus on the user and all else will follow*' (Google, 2018). With the launch of a subsidiary company called "*Ford Autonomous Vehicles LLC*", the Ford Motor Company is self-reportedly embedding a deeper product-line focus where '*the effort is anchored on human-centered design*' (Ford, 2018).

5.2.6. Solution Area (6): Support the role from the triadic ecological paradigm

Theme 6 pertaining to leveraging and augmenting information in the environment and task itself (e.g., situated, ecological, extended cognition, etc.) is expected to gain traction commensurate with technological progress of increased access to ambient data that might have been previously too cost-prohibitive in previous decades. For example, more recent times have seen an acceleration of accessibility from the miniaturization of recording equipment and availability of ubiquitous sensing and computing power. As automation applications continue to grow into new operational areas and expand beyond closed control system process considerations (especially as with vehicles which by definition move from one place to another), recognition of environmental and task dependencies are also expected to grow.

5.3. Limitations

The presently proposed framework to group answers to the potential problems of degraded driver engagement while monitoring driving automation were not derived from a formal and systematic procedure. Instead, the themes were construed in an abductive reasoning manner while trying to organize and relate timely operational concerns (monitoring responsibilities in SAE Level 2 driving automation) with both established and more recently emergent research literature. Assimilation of these solution areas was desirable, considering the long-standing history of general vigilance issues of prolonged human supervisory attention over any automated processes. However, such a framework cannot claim to be the only one conceivable, and the identified themes could be argued to reflect only idiosyncratic knowledge, reasoning, and partial/imperfect readings of a more full body of literature. For example, Themes 1 and 3 were scarcely used categorizations by any of the raters within the present literature survey. Besides clear challenges presented by such a small sample size of only 34 publications, other explanations are also available as to the absence of Themes 1 and 3 among the rater responses. As foreshadowed first by Billings (1991) and repeated by Endsley and Kiris (1995), the rapid release and continual roll-out of automation (then for aviation, now for automotive applications) might obviate a so-called '*too academic*' position of strict avoidance (i.e., Theme 1). Thus, it is conceivable how an approach area as Theme 1 might be under-represented in the literature as being both either too obvious and/or too obsolete. For example, the proactive literature search terms (e.g. of keeping engagement/attention in supervisory control) might reasonably not be expected to return publications that are predominately oriented towards the first solution area of avoiding the supervisory role. In contrast, Theme 3 might be too abstract or unusual (or even arguably unethical as a feature of deception) to be directly arrived at and associated with the terms of '*supervisory control*'. While shared control

and backup automation are far from being alien concepts, the logical complement of changing a subjective experience with automation (Theme 3) to that of changing an objective amount of automation (Theme 2) might be for some too unfamiliar as a grouping umbrella perspective. Furthermore, because humans are still humans whether supervising automated processes or performing other kinds of vigilance and/or sustained attention work, it should be noted that, although presently left out of scope, many of the other literature search returns regarding proactive solutions to human attention/engagement in supervisory or monitoring control/work might be expected to transfer interesting lessons learned even if from non-operator domains: educational classrooms, business offices, creative work, medical hospitals, geriatric care, etc.

6. Conclusions

A wealth of literature suggests categorical approaches to proactive strategies for addressing potential degradation of driver monitoring performance in human supervisory control of driving automation. A qualitative framework of six themes to group solutions have been presently proposed in order to answer a research question of '*how do we keep people engaged while supervising (driving) automation*'. These themes were motivated from human factors and psychological learning theory literature and found to be recognizably applied by raters to categorize empirically grounded human automation interaction research recommendations. The present themes were devised as short-hand formulations that might be easy to remember. Such abstracted organization frameworks are expected to be useful in order to more easily draw comparisons both within and across domains. For example, as a sort of lay of the land overview, the solution areas might serve like a map for automation research/design practitioners to locate where their present approaches (i.e., to human vigilance in supervising driving automation) currently reside and what other alternative areas might be interesting to explore. Additionally, underlying concepts can also thus be more easily entertained to provide common groundwork benefits across seemingly disparate themes.

6.1. General Lessons Learned

The body of literature has much to say regarding supervisory control of automation. We encourage readers towards broader review work in general (Sheridan, 1992), for unmanned robot-vehicle systems (Chen et al., 2011), and for evolving driving roles specifically (Merat & Lee, 2012). Across these review works (and across the six presently identified themes), a consensus benefit would appear to be meta-information requirements to combat uncertainty regarding human involvement in supervising automation (e.g., information about control utility, situated automation capability, performance predictions, etc.). Specific findings from these publications are highlighted below to substantiate this position.

Sheridan (1992) provides a definitive reference for supervisory control that brings together a variety of theories and technologies across decades of his experimental research within the area. In his concluding chapter, he warns of alienation of operators from their work/responsibilities as an underlying cause and concern to be combatted through designs that allow an operator to retain her/her sense of responsibility and accountability. He considers the future of supervisory control in relation to the task entropy (i.e., the complexity or unpredictability of task situations to be dealt with). He offers a way forward through an assumption that humans know best when the automation should apply based on how readily the required information can be modelled.

'The human decision maker is necessary for the information that is not explicitly modelable ... Some, perhaps most, decision situations the human operator will encounter require only information that is modelable. She will make mistakes in such decisions, and can benefit from a decision aid for these cases, and in such cases the decision aid can be validated ... Assume the human can properly decide when the situation includes elements the decision aid can properly assess, and for which elements the decision aid should be ignored' (p. 359).

Chen et al. (2011) cover a multitude of related research concerning human performance issues (e.g., multitasking performance, trust in automation, situation awareness, and operator workload) and innovative technologies designed to reduce potential performance degradations surrounding human supervisory control of automated robot-vehicles. They review interface/tool design developments of multimodal display/controls, planning, visualization, attention management, trust calibration, adaptive automation, and intelligent agent and human-robot teaming. Chen et al. (2011) relay sub-roles within supervisory tasks from Sheridan (2002) that append aspects of planning and learning to bookend monitoring and intervening. Such surrounding aspects of gaining experience with when/where to moderate attention strategies in the application of supervisory control echoes those discussed above by Sheridan (1992).

Complicating interactive challenges reviewed by Chen et al. (2011) include inaccuracies in meta-knowledge that contribute to issues of both automation disuse and over-reliance. On the one hand, humans commonly overestimate the cognitive/perceptual abilities of themselves and others (e.g., metacognitive errors such as change blindness, verbal and visual hindsight bias, self-confirmation bias, cognitive dissonance, etc.) which inflate their sense of necessity for human involvement. On the other hand, to the extent that operators anthropomorphize hardware/software into human-like teammates could then likewise exacerbate expectations of capability, encourage complacency and produce over-reliance on automated processes. At the heart of the issue is the concept of trust calibration where '*operators intervene only when they have reason to believe their own decisions (od) are superior to the automated system's decisions (ad)*' (Chen et al., 2011, p. 437). Within their review of calibrating human trust of automation, Chen et al. (2011) suggest from Lee and See (2004) that '*the capabilities and limitations of the automated systems be conveyed to the operator, when feasible*' because previous research has shown that '*when operators were aware of the context-related nature of automation reliability, their detection rate of automation failures increased significantly*' (e.g., Bagheri & Jamieson, 2004). Beyond aspects of proneness towards false alarms or misses, they suggest additional dimensions of trust: utility, predictability, and intent.

Merat and Lee (2012) include a review of driver automation interaction research to guide future designs. Their results include identification of two general design philosophies for automation: substitution vs. support. They conclude that assumptions towards substitution are not seamlessly simple to meet and instead argue that successful designs will depend on recognizing and supporting the new roles for drivers. Merat and Lee (2012) provide scenario-based warnings both of conflicting timescales: '*Automation may require drivers to intervene on a scale of milliseconds, but reentering the control loop may take seconds*' (p. 683), as well as of ironies of automation that '*...can accommodate the least demanding driving situations—encouraging drivers to disengage from driving—but then calls on the driver to address the most difficult situations ... Periods when drivers are most likely to fully rely on automation—highway driving—also require the most rapid re-*

entry of drivers into the control loop.' (p. 683-684). In consideration of such scenarios, it becomes apparent that interactive meta-information (of humans, vehicles/automation, and the driving task environments) would be essential for forming expectations of how well drivers will perform their monitoring duties.

In summary, a general lesson for common benefit to all solution areas would appear to be further characterizations of driving situations towards understanding which are more complex from those that are more routine (i.e., for both humans and for machines). Such kind of information would support designers and end-user expectations in meta-supervisory mental model knowledge of when/where the automation they are tasked with supervising might better/worse perform and why (and likewise for the monitoring performance/requirements of the human supervisor). To the extent that the driving is able to be handled entirely within perfectly formulated sets of rules and logic, then automated processes should excel and consequences for human oversight would reasonably be diminished. On the other hand, to the extent that driving involves complex socio-cultural norms and violations that are not mathematically well-described and highly interactive with un-modelled context dependencies, then human engagement in monitoring becomes more crucial. For example, as relayed by Merat and Lee (2012): '*Even now, the role of the person behind the wheel is often not that of a driver but that of an office worker on a conference call, a mother caring for a child, or a teen connecting with friends* (Hancock, 2017b)'. As more mutually informed tests are conducted of SAE Level 2 driving automation, between laboratory and on-road research and development, such experiences should serve to provide clearer details, specifics, and evidence in place of assumptions. Positive progress towards specific details relevant for human monitoring of driving automation can be recognized from the California Department of Motor Vehicles. The CA DMV has begun to publically share documentation of annual collision and disengagement reports from autonomous vehicle (test) operations within its jurisdiction (California DMV, 2018) — 95 collision reports are available between 2015-2018, and 2308 disengagements for the 2017 reporting period. More than just a requirement to enumerate problems, the disengagement documentation also begins an attempt to standardize a communication of circumstances (e.g., who initiated the disengagement, on what kind of road, with a description of facts causing the disengagement). Future research might make use of such details to further inform targeted studies surrounding the topic of human attention in supervision of driving automation. As more information becomes available, such information can be used in line with the first three of our presently identified solution area themes to avoid (1) and/or reduce (2-3) the operational design domains of partial automation that requires human supervision, or by the last three solution area themes to support its operations via e.g., enhanced training (4), feedback and mental models (5), and/or task environment relations (6).

Acknowledgments

The research presented in this paper was conducted within the project HFAuto – Human Factors of Automated Driving (PITN-GA-2013-605817). The authors would like to express their gratitude to Rebecca Matthias for her initial conceptualization and guidance regarding the review topic of proactive (vs. reactive) approaches to engagement while supervising driving automation.

References

Note: References marked with a hash # indicate publications included in the present literature survey categorization analysis.

#Ackerman, W., & Block, W. (1992). Understanding supervisory systems. *Computer Applications in Power*, 5(4), pp. 37-40.

Ayre, J. (2017). Ford: Skip level 3 autonomous cars – Even engineers supervising self-driving vehicle testing lose “Situational Awareness”. *Clean Technica*. Available at: <https://cleantechica.com/2017/02/20/ford-skip-level-3-autonomous-cars-even-engineers-supervising-self-driving-vehicle-testing-lose-situational-awareness/>

Bagheri, N., & Jamieson, G.A. (2004). The impact of context-related reliability on automation failure detection and scanning behavior. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pp. 212-217

Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779. [http://dx.doi.org/10.1016/0005-1098\(83\)90046-8](http://dx.doi.org/10.1016/0005-1098(83)90046-8)

Banks, V.A., Eriksson, A., O'Donoghue, J., & Stanton, N.A. (2018). Is partially automated driving a bad idea? Observations from an on-road study. *Applied Ergonomics*, 68, pp. 138-145.

#Banks, V.A., & Stanton, N.A. (2016). Keep the driver in control: Automating automobiles of the future. *Ergonomics*, 53, pp. 389-385.

Banks, V.A., Stanton, N.A., & Harvey, C. (2014). Sub-systems on the road to vehicle automation: Hands and feet free but not ‘mind’ free driving. *Safety Science*, 62, pp. 505-514.

Beggiato, M., Hartwich, F., Schleinitz, K., Krems, J.F., Othersen, I., & Petermann-Stock, I. (2015). What would drivers like to know during automated driving? Information needs at different levels of automation. *7th conference on driver assistance*. Munich, Germany.

Bennett, K.B., & Flach, J.M. (2011). *Display and interface design: Subtle science, exact art*. CRC Press, Taylor & Francis Group. Boca Raton, FL, USA.

Billings, C.E. (1991). Human-centered aircraft automation; A concept and guidelines (NASA Tech. Memorandum 103885). Moffet Field, CA: NASA Ames Research Center.

Black Tesla (2016). Tesla v8.0 Autopilot – Warning Interval & Autosteer Unavailable. Available at: https://www.youtube.com/watch?v=isZ3fSbE_pg

#Blasch, E.P., & Plano, S. (2002). JDL level 5 fusion model ‘user refinement’ issues and applications in group tracking. *Signal Processing, Sensor Fusion, and Target Recognition XI*, pp. 270-290.

BMW (2017). 7 series, 750i xDrive Owner’s Manual. Steering and lane control assistant. Available at: <https://www.bmwusa.com/owners-manuals.html>

#Breda, L.V. (2012). Supervisory control of multiple uninhabited systems: Methodologies and enabling human-robot interface technologies. NATO Research and Technology Organization. Technical Report. Neuilly-Sur-Seine, France.

Broadbent, D.E., & Gregory, M. (1965). Effects of noise and signal rate upon vigilance analysed by means of decision theory. *Human Factors*, 7(2), pp. 155-162.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, 47, pp. 139-159.

Cabral, C.D.D., Goncalves, J., Morando, A., Sassman, M., de Winter, J.C.F., & Happee, R. (2017). HFAuto D3.1 Driver-state monitor. Human Factors of Automated Driving, Initial Training Networks, FP7-PEOPLE-2013-ITN, Grant agreement no.: 605817

Cabral, C.D.D., Happee, R., & de Winter, J.C.F. (2016). From Mackworth’s clock to the open road: A literature review on driver vigilance task operationalization, *Transportation Research Part F: Traffic Psychology and Behaviour*, 40, 169-189.

California DMV (2018). Testing of Autonomous Vehicles with a Driver. State of California Department of Motor Vehicles, Vehicle Registration. <https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/testing>

Card, S.K., Moran, T.P., & Newell, A. (1983). *The psychology of human-computer interaction*. Lawrence Erlbaum Associates. Hillsdale, New Jersey.

Casner, S.M., Hutchins, E.L., & Norman, D. (2016). The challenges of partially automated driving. *Communications of the Association for Computing Machinery (ACM)* , 59(5), pp. 70-77

#Chen, J.Y.C., & Barnes, M.J. (2012a). Supervisory control of multiple robots: Effects of imperfect automation and individual differences. *Human Factors*, 54(2), pp. 157-174.

#Chen, J.Y.C., & Barnes, M.J. (2012b). Supervisory control of multiple robots in dynamic tasking

environments. *Ergonomics*, 55, pp. 1043-1058.

Chen, J.Y.C., Barnes, M.J., & Harper-Sciarini, M. (2011). Supervisory control of multiple robots: Human performance issues and user-interface designs. *IEEE Transactions on Systems, Man, and Cybernetics- Part C: Applications and reviews.*, 41(4), pp. 435-454

Chiappe, D., Strybel, T.Z., & Vu, K.L. (2015). A situated approach to the understanding of dynamic systems. *Journal of Cognitive Engineering and Decision Making*, 9(1), pp. 33-43.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58, pp. 10-23.

#Clauss, S., & Schulte, A. (2014). Implications for operator interactions in an agent supervisory control relationship. *International Conference on Unmanned Aircraft Systems (ICUAS)*. Orlando, FL, USA.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, pp. 213-220.

Craig, A. (1984). Human engineering: The control of vigilance. In J. Warm (Ed.), *Sustained attention in human performance* (pp. 247-291). Chichester, UK:Wiley.

#Cummings, M.L., Bruni, S., Mercier, S., & Mitchell, P.J. (2007). Automation architecture for single operator multiple UAV command and control. *International Command and Control Journal*, 1(2), pp. 1-24.

#Cummings, M.L., & Guerlain, S. (2007). Developing operator capacity estimates for supervisory control of autonomous vehicles. *Human Factors*, 49(1), pp. 1-15.

#Cummings, M.L., Mastracchio, C., Thornburg, K., & Mkrtchyan, A. (2013). Boredom and distraction in multiple unmanned vehicle supervisory control. *Interacting with Computers*, 25(1), pp. 34-47.

#Cummings, M.L., & Mitchell, P.J. (2007). Operator scheduling strategies in supervisory control of multiple UAVs. *Aerospace Science and Technology*, 11(4), pp. 339-348.

#Cummings, M.L., & Nehme, C.E. (2010). Modeling the impact of workload in network centric supervisory control settings. In S. Kornguth, R. Steinberg, & M.D. Matthews (Eds.) *Neurocognitive and physiological factors during high-tempo operations*. Ashgate. Surrey, UK. pp. 23-40.

DARPA (2014). The DARPA Grand Challenge: Ten Years Later. Defense Advanced Research Projects Agency, News and Events. Available at: <https://www.darpa.mil/news-events/2014-03-13>

De Waard, D., Van der Hulst, M., Hoedemaeker, M., & Brookhuis, K.A. (1999). Driver behavior in an emergency situation in the automated highway system. *Transportation Human Factors*, 1(1), pp. 67-82.

De Winter, J.C.F., & Dodou, D. (2011). Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technology and Work*, 16(1), pp. 1-11.

De Winter, J.C.F., Zadpoor, A.A. & Dodou, D.. 2014. "The Expansion of Google Scholar versus Web of Science: A Longitudinal Study." *Scientometrics* 98: 1547-1565

Diewald, S., Moller, A., Roalter, L., Stockinger, T., & Kranz, M. (2013). Gameful design in the automotive domain – Review, outlook, and challenges. *Proceedings of the 5th International Conference on Automotive UI*, Oct. 28-30, Eindhoven, The Netherlands.

Ellingwood, K. (1996). Project begins to test 'driver-less' freeway system. Los Angeles Times. Available at: http://articles.latimes.com/1996-06-28/news/mn-19436_1_highway-system

Elliot, M.A., McColl, V.A., & Kennedy, J.V. (2003) Road design measures to reduce drivers' speed via "psychological" processes: A literature review. (No. TRL Report TRL564). Crowthorne, UK: TRL Limited.

Endsley, M.R. (1995). Toward a theory of situational awareness in dynamic systems. *Human Factors*, 37(1), pp. 32-64.

#Endsley, M.R., & Kiris, E.O. (1995) The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37 (2), pp. 381-394.

Endsley, M.R., Bolte, B., & Jone, D.G. (2003). *Designing for situation awareness: An Approach to User Centered Design*. CRC, Taylor & Francis Group. Boca Raton, FL, USA.

*Eriksson, A., *Petermeijer, S.M., Zimmerman, M., De Winter, J.C.F., Bengler, K.J., & Stanton, N. (in press). Rolling out the red (and green) carpet: Supporting driver decision making in automation to manual transitions. *IEEE Transactions on Human-Machine Systems*. THMS-17-02-0061. *- Joint first author

Fitts PM (ed) (1951) Human engineering for an effective air navigation and traffic control system. National Research Council, Washington, DC.

Flach, J.M. (1990). Control with an eye for perception: Precursors to an active psychophysics. *Ecological psychology*, 2(2), pp. 83-111.

Flach, J.M. (2018). Situated cognition. Professional communication via LinkedIn posting, June 2018. Available at: <https://www.linkedin.com/feed/update/urn:li:activity:6413792059702022144>

Flemisch, F., Abbink, D., Itoh, M., Pacaux-Lemoine, M-P., & Wessel, G. (2016). Shared control is the sharp end of cooperation: Towards a common framework of joint action, shared control and human machine cooperation. *International Federation of Automatic Control (IFAC) PapersOnLine*, 49(19), pp. 72-77.

Ford (2018). Ford creates 'Ford Autonomous Vehicles LLC': Strengthens global organization to accelerate progress, improve fitness. Ford Media Center. Available at: <https://media.ford.com/content/fordmedia/fna/us/en/news/2018/07/24/ford-creates-ford-autonomous-vehicles-llc.html>

Frankmann, J., & Adams, J. (1962). Theories of vigilance. *Psychological Bulletin*, 59(4), 257–272. <http://dx.doi.org/10.1037/h0046142>.

Gallagher, S. (2005). *How the body shapes the mind*. Oxford University Press. Oxford.

Gao, P., Hensley, R., & Zielke, A. (2014). A roadmap to the future for the auto industry. *McKinsey Quarterly*, October.

#Gausshell, D.J., & Darlington, H.T. (1987). Supervisory control and data acquisition. *Proceedings of the IEEE*, 75(12), pp. 1645-1658

#Gersh, J.R., & Hamill, B.W. (1992). Cognitive engineering of rule-based supervisory control systems: Effects of concurrent automation. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*. Chicago, IL, USA. pp. 1208-1213.

Gibson, J.J. (1979). *The ecological approach to visual perception*. Houghton Mifflin. Boston, MA, USA.

Google (2018). Ten things we know to be true. Available at: <https://www.google.com/about/philosophy.html>

Gopher, D. (1991). The skill of attention control: Acquisition and execution of attention strategies. In D. Meyer and S. Kornblum (Eds.), *Attention and Performance XIV*, Hillsdale NJ: Erlbaum.

Greenlee, E.T., DeLucia, P.R., & Newton, D.C. (2018). Driver vigilance in automated vehicles: Hazard detection failures are a matter of time. *Human Factors*, 60(4), p. 465-476.

Groh, F. (2012). Gamification: State of the art definition and utilization. *Proceedings of the 4th Seminar on Research Trends in Media Informatics*, Ulm University, Germany, pp. 39-46.

Gunn, D.V., Warm, J.S., Nelson, W.T., Bolia, R.S., Schumsky, D.A., & Corcoran, K.J. (2005). Target acquisition with UAVs: Vigilance displays and advanced cuing interfaces. *Human Factors*, 47(3), pp. 488-497.

Hancock, P.A. (2017a). On the nature of vigilance. *Human Factors*, 59(1), pp. 35-43.

Hancock, P.A. (2017b). Driven to distraction and back again. In M.A. Regan, J.D. Lee, & T.W. Victor (Eds.). *Driver Distraction and Inattention: Advances in Research and Countermeasures*. Surrey, UK: Ashgate.

#Hart, C.S. (2010). Assessing the impact of low workload in supervisory control of networked unmanned vehicles. Master's Thesis, The Massachusetts Institute of Technology. Available at: <http://web.mit.edu/aeroastro/labs/halab/papers/Hart-Thesis-2010.pdf>

Hawkins, A.J. (2018). Senior citizens will lead the self-driving revolution. The Verge: Transportation, Autonomous Cars. Available at: <https://www.theverge.com/2018/1/10/16874410/voyage-self-driving-cars-villages-florida-retirement-communities>

Hawley, J.K., Mares, A.L., Giannanco, C.A. (2005). The human side of automation: Lessons for air defense command and control. Army Research Laboratory, Report No. ARL-TR-3468.

#Hawley, J.K., Mares, A.L., Giannanco, C.A. (2006). Training for effective human supervisory control of air and missile defense systems. Army Research Laboratory, Report No. ARL-TR-3765.

#Hoc, J.M. (2000). From human-machine interaction to human-machine cooperation. *Ergonomics*, 43, pp. 833-843.

Hutchins, E., Weibel, N., Emmenegger, C., Fouse, A., & Holder, B. (2013). An integrative approach to understanding flight crew activity. *Journal of Cognitive Engineering and Decision Making*, 7(4), pp. 353-376.

Janssen, N.M. (2016). Adaptive automation: Automatically (dis)engaging automation during visually distracted driving. Master's Thesis. Retrieved from TU Delft repository: <http://resolver.tudelft.nl/uuid:dd4cbcc5-f99e-4986-9f82-ab15c621ddc5>

Johnson, J. (2010). *Designing with the mind in mind: Simple guide to understanding user interface design rules*. Morgan Kauffman Publishers, Elsevier. Burlington, MA, USA.

#Johnson, R., Leen, M., & Goldberg, D. (2007). Testing adaptive levels of automation (ALOA) for UAV supervisory control (AFRL-HE-WP-TR-2007-00682). Wright-Patterson Air Force Base, OH: Air Force Research Laboratory.

Kircher, K., & Ahlstrom, C. (2017). Minimum Required Attention: A human-centered approach to driver inattention. *Human Factors*, 59(3), pp. 471-484.

Kyriakidis, M., van de Weijer, C., van Arem, B., & Happee, R. (2015). The deployment of advanced driver assistance systems in Europe. *22nd Intelligent Transportation Systems World Congress*. Bordeaux, France.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, pp. 159-174.

Lee, J.D., & See, K.A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1),

pp. 50-80.

Lee, J.D., & Seppelt, B.D. (2009). Human factors in automation design. In S.Y. Nof (Ed.) *Handbook of Automation*. Springer. Berlin, Heidelberg. pp. 417-436.

Li, L., Wen, D., Zheng, N-N., & Shen, L-C. (2012). Cognitive cars: A new frontier for ADAS research. *IEEE Transactions on Intelligent Transportation Systems*, 13(1), pp. 395-407.

#Lockhart, J.M., Strub, M.H., Hawley, J.K., & Tapia, L.A. (1993). Automation and supervisory control: A perspective on human performance, training, and performance aiding. *Proceedings of the 37th Annual Meeting of the Human Factors and Ergonomics Society*. Seattle, WA, USA.

Lowry, R. (2018). Kappa as a measure of concordance in categorical sorting. VassarStats: Website for Statistical Computation. Available at: <http://vassarstats.net/kappa.html>

Lutteken, N., Zimmermann, M., & Bengler, K. (2016). Using gamification to motivate human cooperation in a lane-change scenario. *IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Rio de Janeiro, Brazil, Nov 1-4.

#Manly, T., Robertson, I.H., Galloway, M., & Hawkins, K. (1999). The absent mind: Further investigations of sustained attention to response. *Neuropsychologia*, 37, pp. 661-670.

Mackworth, N.H. (1948) The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1(1), 6-21, DOI: 10.1080/17470214808416738.

Mackworth, N.H. (1950). Researches on the measurement of human performance. Medical Research Council Special Report Series No. 268. London, UK: HMSO.

Merat, N., Jamson, A.H., Lai, F.C.H., Daly, M., & Carsten, O.M.J. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F*, 27, pp. 274-282.

Merat, N., & Lee, J.D. (2012). Preface to the special section on human factors and automation in vehicles: Designing highly automated vehicles with the driver in mind. *Human Factors*, 54(5), pp. 681-686.

MercBenzKing (2016). 2017 Mercedes S Class Head up display Drive Pilot Nightvision assist Distronic Plus Lane Keeping. YouTube demonstration video. Available at: <https://www.youtube.com/watch?v=RjvI57BIDp0>

Mercedes-Benz (2017). S-Class Operator's Manual. Part no. 2225849405. Available at: https://www.mbusa.com/mercedes/service_and_parts/owners_manuals#!year=2017&class=S-Sedan

#Miller, C., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors*, 49(1), pp. 57-75.

Molloy, R. & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors*, 38(2), p. 311-322.

Moray, N. (1982). Subjective mental workload. *Human Factors*, 24(1), pp. 25-40.

Mosier, K.L., Fischer, U., Morrow, D., Feigh, K.M., Durso, F.T., Sullivan, K., & Pop, V. (2013). Automation, task, and context features: Impacts on pilots' judgments of human-automation interaction. *Journal of Cognitive Engineering and Decision Making*, 7(4), pp. 377-399.

Mulder, M., Abbink, D., & Boer, E. (2012). Sharing control with haptics: Seamless drive support from manual to automatic control. *Human Factors*, 54(5), pp. 786-798.

Neisser, U. (1978). Perceiving, anticipating, and imagining. In W. Savage (Ed.). *Perception and cognition: issues in the foundations of psychology* (pp. 89 – 105). Minneapolis: University of Minnesota Press.

Nelson, J.T., McKinley, R.A., Golob, E.J., Warm, J.S., & Parasuraman, R. (2014). Enhancing vigilance in operators with prefrontal cortex transcranial direct current stimulation (tDCS). *NeuroImage*, 85, pp. 909-917.

NHTSA (2008). National motor vehicle crash causation survey. Report to Congress. *Report No. DOT HS 811059*. Washington, D.C.: National Highway Traffic Safety Administration.

NHTSA (2017). Automated driving systems 2.0: A vision for safety. Washington, D.C.: National Highway Traffic Safety Administration.

Norman, D. (2007). *The Design of Future Things*. New York: Basic books. See esp. chap 3, 'Natural Interaction', section 'Natural Safety', pp. 77-85

#Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R.J. Davidson, G.E. Schwartz, & D. Shapiro (Eds.) *Consciousness and self-regulation (vol. 4): Advances in Research and Theory*. Plenum Press. New York. pp. 1-18.

Ogden, G.D., Levine, J.M., & Eisner, E.J. (1979). Measurement of workload by secondary tasks. *Human Factors*, 21(5), pp. 529-548.

Olson, W.A., & Wuennenberg, M.G. (1984). Autonomy based human-vehicle interface standards for remotely operated aircraft. *Proceedings of the 20th Digital Avionics Systems Conference*, Daytona Beach, FL, USA.

O'Regan, J.K. (1992). Solving the 'real' mysteries of visual perception: The world as an outside memory.

Canadian Journal of Psychology, 46, pp. 461-488.

Panou, M.C., Bekiaris, E.D., & Touliou, A.A. (2010). ADAS module in driving simulation for training young drivers. *13th International IEEE Conference on Intelligent Transportation Systems*. Madeira Island, Portugal.

Parasuraman, R. (1979). Memory load and event rate control sensitivity decrements in sustained attention. *Science*, 205(4409), pp. 924-927.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), pp. 230-253.

Parasuraman, R., Kidwell, B., Olmstead, R., Lin, M.K., Jankord, R., & Greenwood, P. (2014). Interactive effects of the COMT gene and training on individual differences in supervisory control of unmanned vehicles. *Human Factors*, 56 (4), pp. 760-771.

#Parasuraman, R., Galster, S., & Miller, C.A. (2003). Human control of multiple robots in the RoboFlag simulation environment. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*. Washington, DC.

#Parasuraman, R., Galster, S., Squire, P., Furukawa, H., & Miller, C.A. (2005). A flexible delegation interface enhances system performance in human supervision of multiple autonomous robots: Empirical studies with RoboFlag. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 35, pp. 481-493.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, 38 (4), pp. 665-679.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics*, 30(3), pp. 286-297.

Pijnenburg, J. (2017). Naturalism: Effects of an intuitive augmented reality interface property in the display of automated driving status. Master's Thesis. Available at: <http://resolver.tudelft.nl/uuid:92d6b140-56f7-440a-b9ba-ff997c00ab4b>

#Pop, V.L., Stearman, E.J., Kazi, S., & Durso, F.T. (2012). Using engagement to negate vigilance decrements in the NextGen environment. *International Journal of Human-Computer Interaction*, 28, pp. 99-106.

#Pope, A.T., Bogart, E.H., & Bartolome, D.S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, 40, pp. 187-195.

Posner, M.I. (1978). *Chronometric explorations of mind*. Lawrence Erlbaum. Hillsdale, NJ, USA.

Price, M.A., Venkatraman, V., Gibson, M., Lee, J., & Mutlu, B. (2016). Psychophysics of trust in vehicle control algorithms. SAE Technical Paper 2016-01-0144, doi: 10.4271/2016-01-0144.

Rasmussen, J., Pejtersen, A.M., & Goodstein, L.P. (1994). *Cognitive systems engineering*. John Wiley & Sons. New York.

Rudin-Brown, C.M., & Parker, H.A. (2004). Behavioural adaptation to adaptive cruise control (ACC): Implications for preventative strategies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 7(2), pp. 59-76.

#Ruff, H.A., Narayanan, S., & Draper, M.H. (2002). Human interaction with levels of automation and decision aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence*, 11, pp. 335-351.

SAE (2014). Surface vehicle information report: Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. Standard J3016. On-Road Automated Vehicles Standards Committee, SAE International. Available at: https://www.sae.org/standards/content/j3016_201401/

SAE (2018). Surface vehicle recommended practice: Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Standard J3016. On-Road Automated Vehicles Standards Committee, SAE International. Available at: https://www.sae.org/standards/content/j3016_201806/

Saffarian, M., De Winter, J.C.F., & Happee, R. (2012). Automated driving: Human-factors issues and design solutions. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*. Boston, MA, USA.

#Sarter, N.B., & Woods, D.D. (1992). Mode error in the supervisory control of automated systems. *Proceedings of the Human Factors and Ergonomics Society 36th Annual Meeting*. Atlanta, GA, USA.

#Sarter, N.B., & Woods, D.D. (1993). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors*, 37(1), pp. 5-19.

Scallen, S.F., Hancock, P.A., & Duley, J.A. (1995). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Applied Ergonomics*, 26 (6), pp. 397-403.

Schutte, P. (1999). Completion: An alternative to automation. *Journal of Information Technology Impact*, 1(3), pp. 113-118.

#Scott, S.D., Mercier, S., Cummings, M.L., & Wang, E. (2006). Assisting interruption recovery in supervisory control of multiple UAVs. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. San Francisco, CA, USA.

See, J.E., Howe, S.R., Warm, J.S., & Dember, W.N. (1995). Meta-analysis of the sensitivity decrement in

vigilance. *Psychological Bulletin*, 117(2), pp. 230-249.

Seppelt, B.D., & Lee, J.D. (2007). Making adaptive cruise control (ACC) limits visible. *International Journal of Human-Computer Studies*, 65, pp. 192-205.

Seppelt, B.D., & Victor, T.W. (2016). Potential solutions to human factors challenges in road vehicle automation. In G. Meyer & S. Beiker (Eds.) *Road Vehicle Automation 3, Lecture Notes in Mobility*. Springer International Publishing. Switzerland.

#Shaw, T., Emfield, A., Garcia, A., De Visser, E., Miller, C.A., Parasuraman, R., & Fern, L. (2010). Evaluating the benefits and potential costs of automation delegation for supervisory control of multiple UAVs. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*. San Francisco, CA, USA.

#Sheridan, T.B. (1986). Human supervisory control of robot systems. *Proceedings of the IEEE International Conference on Robotics and Automation*, 2, pp. 808-812.

Sheridan, T.B. (1992). *Telerobotics, automation, and human supervisory control*. MIT press. Cambridge, MA, USA.

Sheridan, T.B., Charny, L., Mendel, M., & Roseborough, J.B. (1986). Supervisory control, mental models, and decision aids. U.S. Office of Naval Research, contract report no. N00014-83-K-0193

#Sheridan, T.B., & Hennessey, R.T. (1984). *Research and modelling of supervisory control behavior*. National Academy Press. Washington, DC.

Sheridan, T.B. (2002). Supervisory control. In Humans and Automation: Systems Design and Research Issues. New York: John Wiley and Santa Monica, CA: Human Factors and Ergonomics Society, p. 115-129

Sherman, D. (2016). 2016 BMW 750i xDrive: Second place: Semi-autonomous cars. Car and Driver. Available at: <https://www.caranddriver.com/features/semi-autonomous-cars-compared-tesla-vs-bmw-mercedes-and-infiniti-feature-2016-bmw-750i-xdrive-page-4>

Stanton, N.A. & Young, M.S. (2005). Driver behaviour with adaptive cruise control. *Ergonomics*, 48(10), pp. 1294-1313.

Stanton, N.A., Young, M.S., Walker, G.H., Turner, H., & Randle, S. (2001). Automating the driver's control tasks. *International Journal of Cognitive Ergonomics*, 5(3), pp. 221-236.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 57(4), pp. 421-457.

Strand, N., Nilsson, J., MariAnne Karlsson, I.C., & Nilsson, L. (2014). Semi-automated versus highly automated driving in critical situations caused by automation failure. *Transportation Research Part F: Traffic Psychology and Behaviour*, 27, pp. 218-228.

Suchman, L.A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press. New York.

Super Cars (2017). What Happens If You Leave Tesla Autopilot on FOREVER Terrible Idea. Available at: <https://www.youtube.com/watch?v=C7xV9rMaiNo>

Szymkowski, S. (2017). Waymo found drivers asleep, so it dumped partial self-driving feature. Motor Authority. Available at: https://www.motorauthority.com/news/1113654_waymo-found-drivers-asleep-so-it-dumped-partial-self-driving-feature

Teichner, W.H. (1974). The detection of a simple visual signal as a function of time of watch. *Human Factors*, 16(4), p. 339-353

Terry (2011). The 'Rattomorphism' of gamification. CGP: Critical Gaming Project, University of Washington. Available at: <https://depts.washington.edu/critgame/wordpress/2011/11/the-rattomorphism-of-gamification/>

Tesla (2017). Model S Owner's Manual, 8.0. Driver Assistance Features. Available at: https://www.tesla.com/sites/default/files/model_s_owners_manual_north_america_en_us.pdf

Treat, J., Tumbas, N., McDonald, S., Shinar, D., Hume, R., Mayer, R., Stanisifer, R., & Castellan, R. (1979). Tri level study of the causes of traffic accidents. *Report No. DOT-HS-034-3-535-77* (TAC).

Vicente, K.J., & Rasmussen, J. (1990). The ecology of human-machine systems II: Mediating "direct perception" in complex work domains. *Ecological Psychology*, 2, pp. 207-249.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (Ed.). *Attention and Performance VIII*. Erlbaum. Hillsdale, New Jersey, USA. pp. 239-257.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & Davies (Eds.). *Varieties of attention*. Academic Press. New York. pp. 63-101.

Wickens, C.D., & Carswell, C.M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, 37(2), pp. 473-494.

Wickens, C.D., & Kessel, C. (1979). The effect of participatory mode and task workload on the detection of dynamic system failures. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-9, 24-34.

Chapter 2.2: Supervisory Engagement with Driving Automation

Wiener, E.L, & Curry, R.E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, 23(10), pp. 995-1011.

Wilson, R.A. (2004). *Boundaries of the mind*. Cambridge University Press. Cambridge.

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), pp. 365-375.

Appendix A. Literature Survey List

Inclusion set of categorised human-automation literature conclusions from search for keeping engagement/attention in supervisory control.

Ref ID	Year	First Author	Title
1	2016	Banks	Keep the driver in control: Automating automobiles of the future
2	2014	Clauss	Implications for operator interactions in an agent supervisory control relationship
3	2013	Cummings	Boredom and distraction in multiple unmanned vehicle supervisory control
4	2012	Breda	Supervisory Control of Multiple Uninhabited Systems-Methodologies and Enabling Human-Robot Interface Technologies (Commande et surveillance de multiples ...)
5	2012	Chen	Supervisory control of multiple robots: Effects of imperfect automation and individual differences
6	2012	Chen	Supervisory control of multiple robots in dynamic tasking environments
7	2012	Pop	Using engagement to negate vigilance decrements in the NextGen environment
8	2010	Cummings	Modeling the impact of workload in network centric supervisory control settings
9	2010	Hart	Assessing the impact of low workload in supervisory control of networked unmanned vehicles
10	2010	Shaw	Evaluating the benefits and potential costs of automation delegation for supervisory control of multiple UAVs
11	2007	Cummings	Operator scheduling strategies in supervisory control of multiple UAVs
12	2007	Cummings	Developing operator capacity estimates for supervisory control of autonomous vehicles
13	2007	Cummings	Automation architecture for single operator-multiple UAV command and control
14	2007	Johnson	Testing adaptive levels of automation (ALOA) for UAV supervisory control
15	2007	Miller	Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control
16	2006	Hawley	Training for effective human supervisory control of air and missile defense systems
17	2006	Scott	Assisting interruption recovery in supervisory control of multiple UAVs
18	2005	Parasuraman	A flexible delegation-type interface enhances system performance in human supervision of multiple robots: Empirical studies with RoboFlag
19	2003	Parasuraman	Human control of multiple robots in the RoboFlag simulation environment
20	2002	Blasch	JDL Level 5 fusion model: user refinement issues and applications in group tracking
21	2002	Ruff	Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles
22	2000	Hoc	From human-machine interaction to human-machine cooperation
23	1999	Manly	The absent mind: further investigations of sustained attention to response
24	1995	Endsley	The out-of-the-loop performance problem and level of control in automation
25	1995	Pope	Biocybernetic system evaluates indices of operator engagement in automated task
26	1995	Sarter	How in the world did we ever get into that mode? Mode error and awareness in supervisory control
27	1993	Lockhart	Automation and supervisory control: A perspective on human performance, training, and performance aiding

Ref ID	Year	First Author	Title
28	1992	Ackerman	Understanding supervisory systems
29	1992	Gersh	Cognitive engineering of rule-based supervisory control systems: Effects of concurrent automation
30	1992	Sarter	Mode error in supervisory control of automated systems
31	1987	Gausshell	Supervisory control and data acquisition
32	1986	Norman	Attention to action: Willed and automatic control of behavior
33	1986	Sheridan	Human supervisory control of robot systems
34	1984	Sheridan	Research and modeling of supervisory control behavior. Report of a workshop

Appendix B. Category Ratings

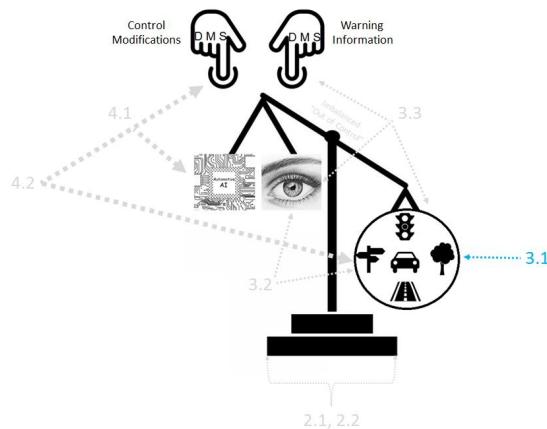
First and second choice (where applicable) thematic category as identified by each rater for each publication reference.
 First choice overlap agreement by at least 2 raters is shaded and full agreement is outlined.

Ref ID	AV 1 st Choice	TX 1 st Choice	CO 1 st Choice	AV 2 nd Choice	TX 2 nd Choice	CO 2 nd Choice
1	5	5	2	2	2	-
2	6	6	2	-	5	6
3	2	2	4	-	5	6
4	6	6	2	5	2	6
5	4	5	4	5	6	-
6	6	5	4	4	2	-
7	4	4	4	-	5	-
8	2	5	2	-	4	-
9	2	5	2	4	2	5
10	2	5	2	-	2	-
11	1	5	2	-	-	-
12	2	2	6	-	5	-
13	5	6	4	6	5	6
14	2	6	2	6	2	6
15	2	2	2	3	6	-
16	4	4	4	-	6	-
17	6	6	6	5	4	5
18	2	5	2	-	6	-
19	6	5	2	-	-	-
20	3	6	6	5	5	-
21	2	5	2	6	-	3
22	5	6	3	6	3	-
23	2	4	4	-	5	-
24	2	1	2	1	2	1
25	2	4	4	-	5	-
26	5	5	6	6	-	4
27	4	4	4	5	6	-
28	5	6	6	6	-	-
29	6	6	3	-	5	-
30	5	5	5	4	-	4
31	6	5	3	5	-	-
32	6	5	2	5	6	-
33	6	5	5	-	6	-
34	5	5	4	4	4	2
Mode:	2	5	2	5	5	6

3

PART 3: Driving Scenes and Driver Eyes

Chap. 3.1) Validity and Reliability of Naturalistic Driving Scene Categorization Judgments from Crowdsourcing



In regards to the overall thesis big picture, this research serves as a direct exploration of the viability of capturing and categorizing driving scenes for applied research at more efficient scales (larger volumes but while retaining satisfactory levels of validity and reliability). An annotation scheme was designed to deliver potentially relevant information about scene contents but without undue burden to human annotators to execute. On average, raters took around 70-75 seconds to complete an annotation of a 3-second driving video clip (e.g., where binary annotation items were pre-sorted by expected frequency likelihoods). By the power of crowdsourcing, 12,892 categorizations were completed in about 1½ days by 200 external workers from 46 different countries. Through volunteer collaboration 1,002 annotation categorizations were completed in about two weeks by six internal confederate workers. The results suggest that large libraries of real-life driving situation visual demands might now be available to generate and organize by recognizable and standardized constituent components. Driving video recording resources could be a real hybrid stimulus boon to driving (vigilance) research such as reviewed in Chap. 2.1 that were found to typically rely on driving environments that are virtual (i.e., simulator studies) or are less controllable/repeatable (i.e., on-road studies). Consequently, an example interface application that allows a researcher to look-up and save a driving video clip by its specified contents is provided in Appendix 3.1.B.2. Video annotations from Chap 3.1 were used to source stimuli for Chap 3.2.

Adapted from:

Cabral, C.D.D., Lu, Z., Kyriakidis, M., Manca, L., Dijksterhuis, C., Happee, R., & de Winter, J.C.F. (2018). Validity and reliability of naturalistic driving scene categorization judgments from crowdsourcing. *Accident Analysis & Prevention*, vol. 114, pgs. 25-33.

Abstract

A common challenge with processing naturalistic driving data is that humans may need to categorize great volumes of recorded visual information. By means of the online platform CrowdFlower, we investigated the potential of crowdsourcing to categorize driving scene features (i.e., presence of other road users, straight road segments, etc.) at greater scale than a single person or a small team of researchers would be capable of. In total, 200 workers from 46 different countries participated in 1.5 days. Validity and reliability were examined, both with and without embedding researcher generated control questions via the CrowdFlower mechanism known as Gold Test Questions (GTQs). By employing GTQs, we found significantly more valid (accurate) and reliable (consistent) identification of driving scene items from external workers. Specifically, at a small scale CrowdFlower Job of 48 three-second video segments, an accuracy (i.e., relative to the ratings of a confederate researcher) of 91% on items was found with GTQs compared to 78% without. A difference in bias was found, where without GTQs, external workers returned more false positives than with GTQs. At a larger scale CrowdFlower Job making exclusive use of GTQs, 12,862 three-second video segments were released for annotation. Infeasible (and self-defeating) to check the accuracy of each at this scale, a random subset of 1,012 categorizations was validated and returned similar levels of accuracy (95%). In the small scale Job, where full video segments were repeated in triplicate, the percentage of unanimous agreement on the items was found significantly more consistent when using GTQs (90%) than without them (65%). Additionally, in the larger scale Job (where a single second of a video segment was overlapped by ratings of three sequentially neighboring segments), a mean unanimity of 94% was obtained with validated-as-correct ratings and 91% with non-validated ratings. Because the video segments overlapped in full for the small scale Job, and in part for the larger scale Job, it should be noted that such reliability reported here may not be directly comparable. Nonetheless, such results are both indicative of high levels of obtained rating reliability. Overall, our results provide compelling evidence for CrowdFlower, via use of GTQs, being able to yield more accurate and consistent crowdsourced categorizations of naturalistic driving scene contents than when used without such a control mechanism. Such annotations in such short periods of time present a potentially powerful resource in driving research and driving automation development.

1. Introduction

Further knowledge specifically of (background) driving scene contexts could benefit transportation research and ultimately road safety. This study presents and evaluates a new method using crowdsourcing to provide content characterizations of natural driving video footage. Brief descriptions of both topics are provided in the following introductory sections.

1.1. **Naturalistic driving and driving videos**

Naturalistic driving studies (NDS) have been growing in popularity with much success over the last few decades. NDS offer advantages with respect to other traditional driving safety research methods such as eye witness recall (often being inaccurate or unavailable) within crash data evidence approaches and driving simulators (often causing artificial participant behavior) (Regan et al., 2012). However, a lack of experimental control (where extraneous variables except that of manipulative interest are held constant), has been a commonly recognized detriment to NDS. Thus, the accurate annotation of the situational aspects and conditional characteristics that freely vary in NDS becomes all the more important for the identification and understanding of potential causal factors. Augmented by accelerating developments in audio-visual technology, computing, and networking resources, blended research designs are emerging wherein stimuli can be naturally sourced from the real world, reproduced, and mixed with more controlled laboratory conditions.

Due to reductions both in size and costs of cameras, real life driving video is an increasingly accessible data resource that may allow recordings at a large scale and could help enrich other sources of data with otherwise missed contextualized information. However, so much video data might be recorded in naturalistic driving research and field operational tests that research resources are often overwhelmed to process such data libraries through pre-requisite rounds of organization and labeling (e.g., data reduction) towards fuller potentials of use. For example, challenges can arise regarding the availability of confederate researchers for laborious manual annotation or transcription tasks. Unfortunately for driving safety research, the use of real-life driving video footage has remained a relatively low-tapped exception (e.g., Crundall, Underwood, & Chapman, 1999; Chapman et al., 2007; Borowsky, Shinar, & Oron-Gilad, 2010) rather than a common resource, despite inherent strengths in face validity and generalizability of results.

1.2 **Crowdsourcing**

Compared to less than 1% in 1995, about 48% of the world population has an Internet connection to date, placing the approximate number of Internet users in excess of 3.5 billion people (www.InternetLiveStats.com/internet-users/). Online crowdsourcing services make use of this extensive connectivity to create an on-call global workforce to complete large projects in small chunks (a.k.a., micro-task workers). Gosling and Mason (2015) review a broad and growing use of Internet resources in recent psychological research. They conclude that harnessing large, diverse, and real-world data sets presents new opportunities that can increase the societal impact of psychological research. In the automated driving domain, research has recently begun to emerge utilizing crowdsourcing resources through global survey initiatives to capture large scale international public opinion (Bazilinskyy & De Winter, 2015; Kyriakidis, Happee, & De Winter, 2015). In regards to crowdsourcing as a research method, investigation into the differences between laboratory participants versus crowdworkers has found faster responses but higher false alarms with crowdsourcing (Smucker & Jethani, 2011). Additional methodological research has revolved

around the assurance of quality from the quick and inexpensive results typically returned by crowdsourcing and have recommended predetermined answer sets for use both in the screening of unethical workers as well as for the effective training of ethical workers (Le et al., 2010; Soleymani & Larson, 2010).

1.3. Present study

Real-world driving datasets come with large labor challenges in terms of data reduction like manual annotation and categorization. Pairing together expansive datasets of naturalistic driving video footage with crowdworkers may be a powerful method for progressing driving safety research. As a prototypical example of the power of crowdsourcing, the online platform known as CrowdFlower can accomplish routine categorization work at relatively low cost and at high speed by distributing the work around the world, taking advantage of both differences in time zones and hourly wages. However, such new methods require an investigation of validity and reliability to ensure trustworthy results might still be retained when scaling up beyond a single researcher or small research team. The present study investigated the use of CrowdFlower in the categorization of large amounts of videos with diverse driving scene contents (i.e., presence of another vehicle, straight road segments, etc.) through manipulation of one of its central quality control mechanisms to ascertain the quality and capability of such a method.

2. Methods

2.1. Quality control settings

Within its documentation, the CrowdFlower system promotes Gold Test Questions (GTQ) as its most important quality control mechanism. By configuring this setting, we enforced that a set of categorizations with known answers (i.e., given by the experimenters) were randomly intermixed with the experimental categorizations of interest. Thresholds of performance on these GTQs were set in an attempt to reduce the amount of indiscriminate responses that may occur within the results due to the remotely distributed nature of work under unsupervised conditions.

2.2. Participants/Workers

Participants in this research consisted of external micro-task workers from the online CrowdFlower contributor community. From this network, workers were prescreened by a number of criteria selectable within the CrowdFlower interface. Specifically, within CrowdFlower, performance levels are automatically awarded based on CrowdFlower's criteria of accuracy across a variety of different Job types. We selected a performance setting of Level 2 workers from a three-level scale, representing the midpoint between anchors of "highest speed" (Level 1) and "highest quality" (Level 3). Moreover, across all 51 of its current possible Channels for sourcing external workers (e.g. BitcoinGet, ClixSense, CoinWorker.com, etc.), CrowdFlower was set to include workers only from those retaining a ratio of Trusted to Untrusted Judgments greater or equal to 80% (39 Channels were left toggled on and 12 set to off). All countries were permitted within the Geography setting, and no additional Language Capability requirements were selected. Table 3.1.1 lists the countries and source Channels of workers obtained across different sets of categorizations performed within the present study along with distributions of unique worker IP addresses and CrowdFlower worker IDs while Fig. 3.1.1 depicts the country distribution of the workers. For external crowdworkers, identification of country was determined by CrowdFlower based on IP address.

Table 3.1.1. Overview of the five different sets of categorizations. These sets included differences in the amount of video segments to be categorized (C1 = 48 segments, C2 = 12,862 segments), the use of Gold Test Questions (C1b had none) and the relation of the annotators to the research (external = CrowdFlower workers; internal = confederate research team).

Condition	Countries (ISO 3166-1 alpha-3)	Channels	Unique IP's	Unique ID's
C1a	15 = AUT, BEL, COL, DEU, ESP, GBR, GRC, IND, MKD, PHL, PRT, ROU, RUS, SRB, TUR	5 = clixsense, coinworker, elite, prodege, tremorgames	18	18
C1b	9 = DNK, GRC, IND, MDA, PAK, PHL, SRB, TUR, VNM	3 = clixsense, elite, tremorgames	13	13
C1c	1 = NLD	1 = n/a (internal)	1	1
C2a	46 = ARG, AUS, AUT, BEL, BGD, BGR, BIH, BRA, CAN, CHL, CZE, DEU, ESP, FIN, FRA, GBR, GRC, HRV, HUN, IDN, IND, ISR, ITA, JAM, LKA, MAR, MDA, MEX, MKD, MYS, PER, PHL, POL, PRT, ROU, RUS, SAU, SRB, SWE, TUR, TWN, UKR, URY, USA, VEN, VNM	16 = clixsense, coinworker, fusioncash, gifthulk, hiving, indivilagetest, instagc, personally, pocketmoneygypt, points2shop, prodege, superrewards, surveymad, tremorgames, yute_jamaica, zoombucks	247	200
C2c	1 = NLD	n/a (internal)	12	7

Note. Country abbreviations are according to ISO 3166-1 alpha-3.

2.3. Apparatus and stimuli

To support projects oriented around the human factors of automated driving (i.e., exposing participants to various HMI/functional research concepts, measuring constructs of vigilance, situation awareness, mental models, reaction time, eye tracking behavior, etc.), a set of stimulus material was desired that had both qualities of high visual realism and controllable levels of uncertainty in repetition, freeze-ability, etc. Initial searches of YouTube with the keyword “dash cam” were conducted to compile a sample database of naturalistic driving video footage. Videos had to feature relatively high and consistent visual quality, a large and consistent field of view, and uninterrupted driving in order to be included. Candidate videos were selected from the search results in order to acquire nominal driving footage (i.e., excluding violations and crashes). We collected a set of 10 freely available YouTube videos ranging between 1 minute and 1 hour duration (but of bimodal typicality of about 3 or 13 minutes length) for a total of 6,934 seconds of driving

footage. The countries in which the recordings were filmed were not known, but driving was always on the right hand side. Audio was removed from the videos.

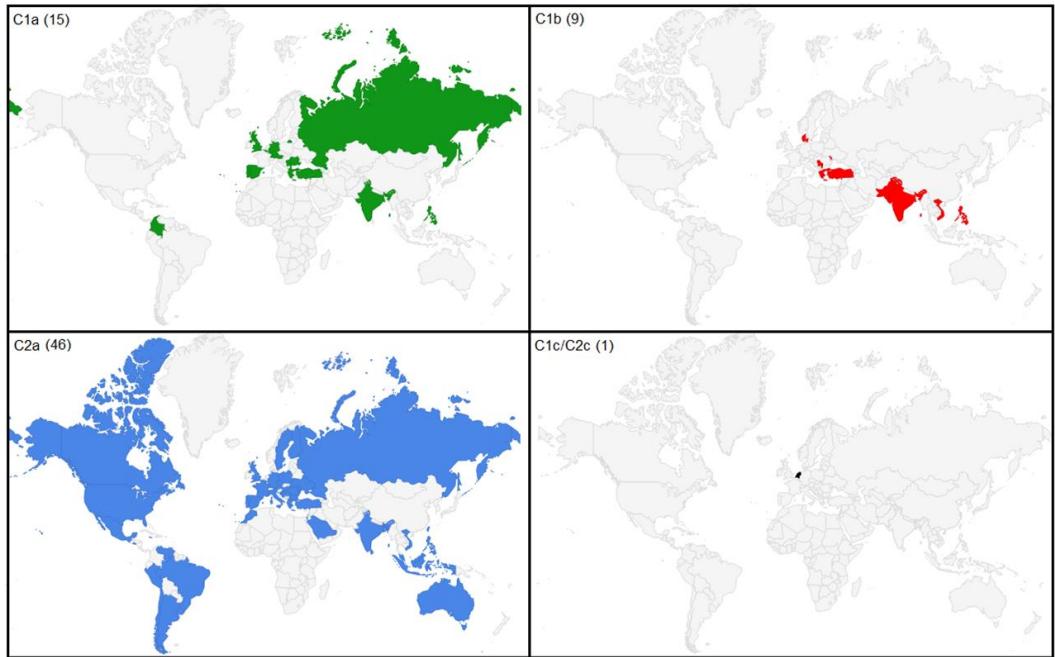


Figure 3.1.1. Annotator country locations by condition.

Subsequently, new self-recorded dash cam driving recordings (6,026 seconds) were filmed in the United States and saved as 39 different files (typically less than 3 minutes in length, but ranging up to 15 minutes). This complemented the videos collected from YouTube in order to exhibit a broader range of real-life and experimentally interesting driving situations. These additional recordings included driving at night, on mostly empty desert roads, in a visually complex metropolis, and via multi-lane freeways, as well as at different driving speeds.

Driving videos from both sources were uploaded as 49 new private link-only access YouTube videos ($M = 264$ seconds duration) with an aggregate of 12,960 seconds of near driver point-of-view video footage. Through a combination of MATLAB script and an online tool from www.tech-tipsforall.com (ttfaloopandrepeat.appspot.com), auto-cueing URL links were generated to access each of the 12,862 possible 3-second segments from each of these 49 video. These URL links were embedded as text only in our CrowdFlower surveys with one URL per Judgment. The video segments overlapped in a manner such that a randomly selected worker categorized seconds one to three from video 1, another randomly selected worker categorized seconds two to four from video 1, a third randomly selected worker categorized seconds three to five from video 1, etc., for all videos 1 through 49. Example screenshots from the driving video segments are shown in Figs. 3.1.2a, 3.1.2b, and 3.1.2c.



Figure 3.1.2. Example screenshots from driving video segments a) recorded from within a publically posted dash cam YouTube video, b) recorded by the experimenters within a visually complex metropolis (i.e., Las Vegas strip), and c) recorded by the experimenters in a visually simple environment (i.e., Nevada desert backroad). Video resolution/quality here is only approximately representative as that initially made available to participants because differences in devices and browsers, full-screen viewing, etc. were not controlled for in the online survey.

A coding scheme was created wherein each video segment categorization (i.e., Judgment) contained two groups of questions. The first group consisted of 21 checkbox items pertaining to the non-mutually exclusive presence of others, namely, (1) cars/trucks/vans/buses, (2) motorcycles/scooters/mopeds, (3) bicycles, and (4) pedestrians. Each of these four categories contained additional possible sub-specification of their position/direction of travel, namely, (5–8) leading, (9–12) oncoming, (13–16) passing or being passed, and (17–20) crossing; all relative to the present point-of-view vehicle. Additionally, there was a checkbox item which should be ticked for (21) no one else was present.

The second group consisted of 10 checkbox items pertaining to presence of miscellaneous infrastructural elements and aspects of vehicle behavior. These were: (1) straight road, (2) more than one lane per direction of travel, (3) signs/signals facing the driver, (4) road surface markings other than lane boundaries (e.g., crosswalks, arrows, writing, etc.), (5) lane change by this driver, (6) lane change by another vehicle, (7) turning by this driver, (8) turning by another vehicle, (9) this driver slowing to a stop, and (10) none of the above. In the second round of categorizations (C2, see Tables 3.1.1 & 3.1.2), the coding scheme was extended to include a position/direction item across all road user categories (i.e., of being parked/stationary), plus a miscellaneous item for overt video edits/alterations. Consequently, these extensions (for further data enrichment value) raised the total checkbox count per video segment to 36. The full coding scheme of annotation items (as well as the specific full training instructions given to annotators) is provided in Appendix A.

2.4. GTQ video segments: multiple purposes and representative examples

GTQ videos were selected from the full pool of video segments under the criteria to serve as effective screening and training devices. For the purpose of screening indiscriminate respondents, some of the easiest and most unambiguous scenes were selected, as for example a video segment where only an empty desert road is shown:

(1) <https://www.youtube.com/embed/eS79DG08idY?start=12&end=15>

For the purpose of explicating various annotation labels (e.g., surface paint markings, signage facing the driver), video segments were selected that contained certain items of interest, such as a segment where a railroad crossing sign appears on the side of the road as well as surface markings in the lane of travel:

(2) <https://www.youtube.com/embed/vA5AiKbzIww?start=82&end=85>

2.5. Conditions

Three different external CrowdFlower Jobs were conducted in two different rounds (C1 and C2), as shown in Table 3.1.2. In the first round, C1, a set of 48 unique three-second long video segments (randomly selected from the larger full dataset of collected video footage) were categorized by external CrowdFlower workers with GTQs either turned on (C1a) or turned off (C1b). In C1a and C1b, the default triplicate redundancy setting in CrowdFlower was kept on and so the Job ran until three Judgments were collected for each video segment. Additionally, the same 48 segments were categorized offline by an individual internal worker (i.e., a confederate researcher) in C1c.

In the second round, C2, Judgments were performed on CrowdFlower across all 12,862 possible 3-second video segments of the full video dataset via external CrowdFlower workers (C2a) and over a subset of these video segments by an internal worker team comprised of multiple confederate researchers (C2c) using the same CrowdFlower structure as the external workers. Within the C2c round of internal team ratings, one team member accomplished a high volume of Judgments ($n = 638$) under two separate CrowdFlower accounts such that 38 different Judgments of the same driving scene segment from the same person were available to establish intra-rater reliability.

The required set of Judgments ordered for each CrowdFlower Job was specified at Job launch and included a redundancy option through a multiplier setting (x3 was used in C1, x1 was used in C2).

Table 3.1.2. Categorization conditions.

Condition	Workers	Video segments categorized	Redundancy	Gold Test Questions	Video segments per Page	Worker payment per Page	Total CrowdFlower Cost
C1a	external	48	3	12	10	\$0.50	\$10.80
C1b	external	48	3	0	10	\$0.50	\$9.00
C1c	internal	48	1	12	n/a	n/a	n/a
C2a	external	12,862	1	53	11	\$0.25	\$349.32
C2c	internal	1,012	1	42	11	n/a	n/a

Note. The total worker payment differs from the total CrowdFlower costs because CrowdFlower retained a margin of about 20%. Video segments per Page refers to the amount of videos the worker was assigned at a time (i.e., stacked vertically, with a scrollbar); total Pages completed varied between workers. A single Page consisted of 10 (C1) or 11 (C2) Judgments, that is, different driving video segments to be annotated.

2.6. Analyses

In the investigation of the utility of CrowdFlower for annotating driving video content, multiple analyses from two different rounds of Jobs (Table 3.1.1) were undertaken to cover the separate but related psychometric aspects of validity (i.e., accuracy) as well as reliability (i.e., consistency).

In terms of validity, we ascertained to what extent categorizations returned from external CrowdFlower workers reflect what is actually visible in a given driving video segment. At an initial reduced Job scale, the same set of video segments was repeated with and without GTQs (Table 3.1.1, C1a vs. C1b) and compared to a reference set of categorizations of these same segments generated by a confederate researcher (C1c). For subsequent accuracy analyses at the greater Job scale (where GTQs were retained), ground truth was created by a team of internal confederates for a random subset due to the infeasibility (and self-defeating purpose) of checking the accuracy of each annotation at this scale.

In terms of reliability, we assessed how consistent categorizations of the driving video segments were when repeatedly administered. Supporting this aim, three analyses were conducted. First, from the second round of confederate categorizations (C2c) one internal team member was given a subset to categorize in duplicate to himself (i.e., randomly intermixed among his other categorizations, see 2.5 Conditions). Second, at the small scale Job (C1), each video segment was rated by three different external CrowdFlower workers (both in C1a and in C1b). Third, the full dataset categorizations of C2a provided an account of consistency due to the fact that the video segments overlapped such that any second of driving video footage was categorized three times. That is, for any second “x” bounded by start/end points [start, end] there existed a first segment: [x, x+2], a second segment: [x-1, x+1], and a third segment: [x-2, x].

2.7. Procedure

All workers were provided with a set of instructions and examples regarding the driving video segment categorization coding scheme that remained available for consultation throughout their work (Appendix A). A single Judgment consisted of a set of 31 (C1) or 36 (C2) checkboxes pertaining to features visible within a randomly selected 3-second long driving video segment (Section 2.3). A single Page consisted of 10 (C1) or 11 (C2) Judgments, that is, different driving video segments to be annotated.

In the conditions where GTQs were active (C1a, C2a, C2c), task workers were first given a single page of Quiz Mode GTQs Judgments to complete. Because of constraints of CrowdFlower, a GTQ Judgment had to be answered perfectly in order to be scored as correct, with no partial credit given (i.e., all 31 or 36 checkboxes had to be checked correctly against predetermined answers constructed by the experimenters). If workers achieved a threshold correctness Trust Score on these GTQs of 70% [i.e., 7 out of 10 Judgements] in C1, and 25% [i.e., 3 out of 11 Judgments] in C2, then workers were automatically allowed by CrowdFlower to continue through as many more Pages of Work Mode as they would like. Through trial and error, the set threshold was lowered from 70% in C1 to 25% in C2, because it turned out to be often highly difficult to obtain a perfect answer on each of the checkboxes of a Judgment. Additionally, in C2, participants were supported with further detailed feedback explaining the correct answers. For an incorrect answer to any checkbox item of a GTQ during Quiz Mode, workers were shown the correct answers of all checkboxes for that Judgment along with a brief justification. Each Page of Work Mode had one new not-yet-seen GTQ randomly presented within the other Judgments such that a worker was unable to identify which Judgments had a priori answers that their own answers would be scored against. As long as workers maintained a running average Trust Score above the set threshold (i.e., 70% in C1, 25% in C2), and there were still GTQs remaining that they had not yet seen, they were allowed to continue.

In the CrowdFlower condition without GTQs (C1b), workers were allowed to enter Work Mode straightaway without real-time screening criteria barring them from submitting Judgments. On a first-come-first-serve (optionally screened) basis, Jobs in CrowdFlower are run until a pre-determined amount of Judgments are completed by an indeterminate amount of workers.

In summary, the GTQ condition included further screening and training to enhance the responses of task workers than the condition without GTQs.

3. Results

The utility of the crowdsourcing platform CrowdFlower in the content categorization of naturalistic driving video footage was investigated through multiple analyses concerning both validity and reliability. Overall, the supposed utility of CrowdFlower in the present tasks was found to be supported (see Table 3.1.3). Results were indicative of significantly increased utility both in terms of validity and reliability in the presence of GTQs as compared to without GTQs. Results were obtained both in the preliminary round of a reduced scale (C1: 48 video segments) and in the subsequent round conducted at a larger scale (C2: 12,862 video segments).

Table 3.1.3. Summary of analyses.

Section	Analysis aim	Relative Job size	Analysis outcome
3.1.1	Validity	Small	The GTQ condition yielded more accurate Judgments than the No GTQs condition. Accuracy was assessed by using the Judgments of a single internal confederate rater as ground truth.
3.1.2	Validity	Large	The GTQ condition yielded accurate Judgments. Accuracy was assessed by using the Judgments of a small team of internal confederate raters as ground truth.

Section	Analysis aim	Relative Job size	Analysis outcome
3.2.1	Reliability	Small	A single confederate rater was found to be consistent to himself.
3.2.2	Reliability	Small	The GTQ condition yielded more consistent Judgments than the No GTQ condition, for full Judgments and at the item level.
3.2.3	Reliability	Large	The GTQ condition yielded Judgments of high inter-rater consistency for overlapping video segments. Consistency was assessed for known-to-be-accurate Judgments.
3.2.4	Reliability	Large	The GTQ condition yielded high inter-rater consistency for overlapping video segments. Consistency was assessed for unknown-to-be-accurate Judgments.

3.1. Validity

3.1.1. 48 Judgments, comparing GTQ with no GTQ

Results showed that there were 35 of 144 (24%) and 6 of 144 (4%) exact matches from C1a (with GTQs) and C1b (without GTQs) respectively, relative to C1c (taken as a measure of ground truth). Results thus indicated inaccuracies in the Judgments from both C1a and C1b (Fig. 3.1.3).

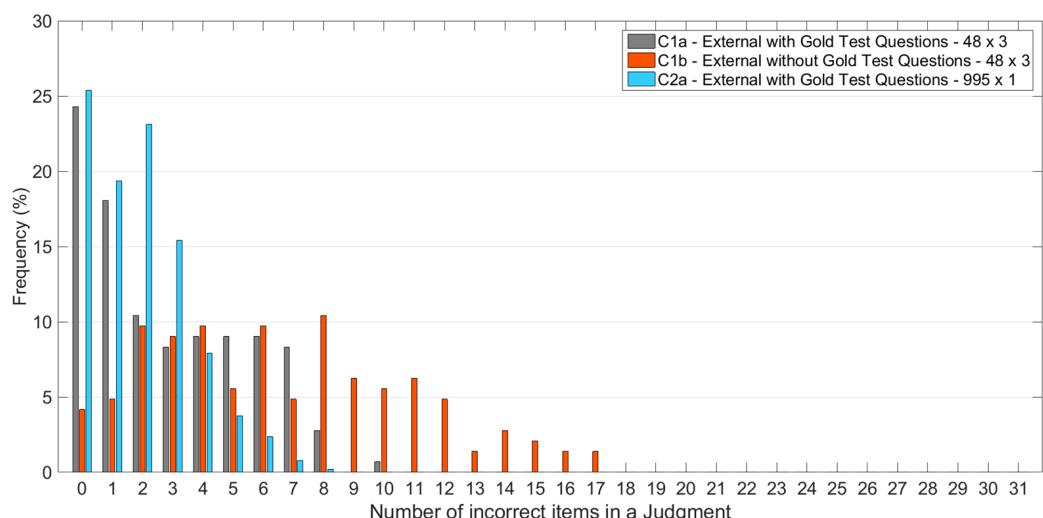


Figure 3.1.3. Distribution of the number of errors per Judgment at the smaller C1 Job scale of 144 Judgments (with and without GTQs) and for a subset of 995 Judgments from the larger C2 Job scale (with GTQs). Errors were determined against known answers (C1c or C2c). A score of 0 signifies a perfectly correct Judgment.

However, these inaccuracies occurred in different specificity/sensitivity biases. Phi correlation coefficients were computed between each full Judgment (i.e., an array of 31 binary checkboxes) from a condition (C1a or C1b) against the ground-truth Judgment returned by an internal confederate rater (C1c) matched for a specific video segment. The median across all 144 (48 x 3) correlation coefficients of the GTQ condition (C1a; $r = 0.78$) was significantly higher than for the No GTQ condition C1b ($r = 0.39$) (Mann-Whitney $U = 3756$, $n_1 = n_2 = 144$, $p < 0.001$ two tailed).

Furthermore, greater total item accuracy across all 4,464 (31 x 48 x 3) categorized items was found in C1a (4,051 = 91%) than in C1b (3,504 = 78%).

Among the 4,464 categorized items in C1b (i.e., without GTQs), there were 396 false positives (i.e., items marked present but which were absent in the video segment according to the confederate researcher), yielding a false positive rate of 11% (396/3,519). Furthermore, there were 564 misses (i.e., items marked absent that were present in the video segment according to the confederate researcher), yielding a miss rate of 60% (564/945). In C1a (with GTQs), the false positive rate was 1.6% (57/3,519) and the miss rate was 38% (356/945). In other words, GTQs contributed to a reduction of both false positives and false negatives.

3.1.2. 1,012 Judgments, comparing external versus internal workers

The confederate research team (C2c) performed 995 Judgments of video segments (17 video segments were removed due to video playback errors) which were randomly selected from C2a. Results showed that there were 257 (26%) exact matches between the Judgments from C2a and C2c. Phi correlations with the ground truth for both the smaller scale Job (correlation between C1a and C1c: median $r = 0.78$, see also Section 3.2.1) and the larger scale Job (correlation between C2a and C2c: median $r = 0.80$) were not found to significantly differ (Mann-Whitney $U = 65298.5$, $n_1 = 144$, $n_2 = 995$, $p = 0.083$).

From the 35,820 C2a items re-rated within C2c (995 Judgments x 36 items per Judgment) the false positive rate was 2.1% (682/31,564) and the miss rate was 27.6% (1,176/4,256).

3.2. Reliability

3.2.1. 38 Judgments, comparing confederate to himself

In condition C2c, one confederate performed 638 Judgments about evenly split under two different CrowdFlower accounts, with an approximate 10% subset of his Judgments from each account coded in duplicate ($n = 38$). Intra-individual test-retest reliability results for this same rater using the same software settings but across different sessions were: 34 (89%) exact matches, an average phi correlation of 0.98 across the 38 Judgments, and an overall item accuracy of 99.5% (i.e., 1,361 out of 1,368).

3.2.2. 48 Judgments, comparing GTQ versus no GTQ

During C1a and C1b, each video segment collected three external worker Judgments and so allowed for a consistency measure of how many categorization ratings (both for full Judgments and/or across items within Judgments) were returned identically between external CrowdFlower task workers. Unanimous agreement on all 31 items of a Judgement was found in 7 of 48 Judgments in C1a (with GTQs) and in 1 of 48 Judgments in C1b (without GTQs). Per item, the unanimous agreement percentage across the 48 Judgments was computed, and was found to be significantly higher for C1a ($M = 90\%$, $SD = 13$) than for C1b ($M = 65\%$, $SD = 19$, $n_1 = n_2 = 31$, $t(60) = 5.85$, $p < 0.001$).

3.2.3. 257 Judgments, comparing ratings by unanimous voting

For the correct 257 Judgments in C2 (see Section 3.1.2), a reliability analysis was conducted by comparing overlapping categorizations across sequential seconds of video footage. For example, the correct true/false answer provided for an item in a video segment that began at time x , was compared with the answer received for that same item by another external worker whose video segment began at time $x-1$ and additionally by another external worker whose video segment began at time $x-2$. It should be noted that some variation between overlapping video segments would be expected to exist (e.g., a car seen only in the last second of a segment that starts at $x = 0$ might not be visible in the previous videos $x-1$ and $x-2$). Due to such uncertainty, somewhat less than perfect reliability may be expected even from perfectly reliable raters. This necessitates consideration of proportional consistency analysis across the entire array of 36 items contained within a Judgment. In other words, it is assumed that while one or a few aspects might vary between overlapping videos, the majority of aspects should remain the same.

Results showed that 74 of 257 correct Judgments (29%) received the same true/false rating across all 36 items by three different external workers who rated overlapping video segments. Figure 3.1.4 shows a distribution of the 257 Judgments according to the number of items yielding unanimous agreement. Judgments always had more than two-thirds (i.e., at least 25 out of 36 items) unanimous agreement, and the mean number of items yielding unanimous agreement was 33.9 out of a possible 36.

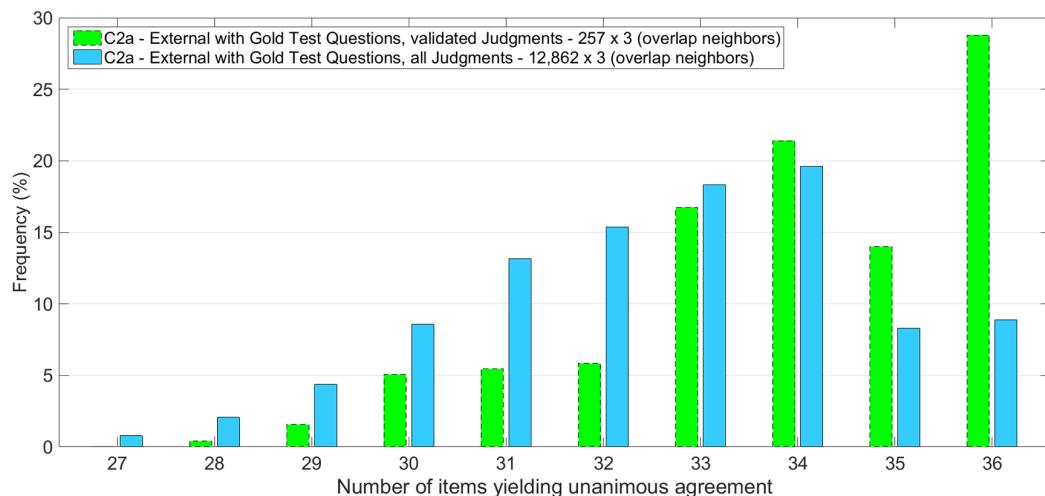


Figure 3.1.4. Frequency of validated (i.e., 257 fully correct) and all returned Judgments (originally 12,862) from C2a according to number of items yielding unanimous agreement from three independent raters.

3.2.4. 12,862 Judgments, comparing ratings by unanimous voting

For all 12,862 Judgments, a reliability analysis of unanimous answers was conducted with overlapping sequential seconds again as in Section 3.2.3, but now for the full dataset. The first and last two Judgments of each video required removal due to a logical lack of full overlap, resulting in a total of 12,670 Judgments (12,862 – 4 x 48).

Regarding unanimity of full Judgments, 1,129 of 12,670 answers (9%) received the same true/false value across all 36 items by the three different external workers. The mean number of items with unanimous agreement per Judgment was 32.6 out of 36 possible.

The distributions of Judgments in Figure 3.1.4 shows that disagreement existed in the categorizations of overlapping sequential seconds of video footage; this occurred most frequently for two items.

4. Discussion, Conclusions and Recommendations

The CrowdFlower crowdsourcing platform may present great potential for driving research by bringing task workers from across the world to categorize a rapidly growing resource of naturalistic driving video data. Due to its inherently distributed structure, CrowdFlower and online tools of similar kind may be more susceptible to fraudulent or non-discriminating responses as compared to locally administered and more tightly controlled traditional methods. Specifically, the utility of CrowdFlower with (and without) its self-purported most important quality control mechanism of GTQs was investigated in the objective categorization of driving video contents via binary presence/absence flagging of pre-specified driving items of interest both at a preliminary reduced and a subsequently increased Job scale.

Exhibiting credible signs of validity and reliability (Table 3.1.3), the potential for the method of crowdsourcing the categorization of driving video contents can be considered in a meaningful and valuable way. For example, as a result of our settings in the present study, 12,862 CrowdFlower annotation categorizations were completed in about one and a half days by 200 external workers from 46 different countries working at an hourly rate of 1.09 USD each (total cost of about 349.32 USD inclusive of a 20% transaction fee) with an average of 75 seconds per Judgment. Through volunteer confederate collaboration, 1,002 annotation categorizations were completed in about two weeks by six internal confederate workers from the Netherlands working between/around their other work duties at a conservative estimated hourly rate around \$20.25 USD each (total cost estimate of about \$394.54 with an average of 70 seconds per Judgment). Thus, for the same approximate costs, the external workers returned categorizations about ten times faster.

Several limitations exist within the present study and are worth mentioning. The first and foremost, is that the GTQ mechanism is explicitly designed to work with objective tasks where there are clear and definable right and wrong answers and so it may not be suitable for many otherwise desirable subjective judgments from a distributed task worker network. A GTQ is constructed in CrowdFlower to require pre-defined correct answers with as minimal ambiguity as possible as well as detailed and documentable justification/motivation of that answer (similar to how both annotator screening and training is used in more controlled laboratory experiments). It should be noted that the design of the present study does not lend itself towards some other research questions that might be addressed from pairing crowdsourcing to naturalistic driving data for example for purposes of investigating the general human ability in perception/annotation of various aspects of driving scenes (inter-item research questions) and/or the bearing of universal/local driving cultures on driving scene interpretation (inter-cultural research questions). Instead, the present study aimed to eliminate ambiguities on an equal par between conditions to test the principle manipulation of interest: the use or not of GTQs.

Nonetheless, some of our requested annotation items appear to have contributed to some confusion between some raters. The worst three annotation items, both in terms of accuracy and reliability, pertained to identification of fully straight roads, signage/signals facing the driver, and number of lanes per direction of travel. Overall, performance with these items averaged around 63% (reliability) and 79% (accuracy) compared to averages taken across all the remaining items of 93% (reliability) and 96% (accuracy). Without proper hypotheses/controls in place, we cannot propose these as particularly systematic nor meaningful results in human perception or suitability to crowdsourcing beyond our own inabilities to more thoroughly formulate such desired details for our driving video data library into more fully objective definitions/terms (see Appendix A). For example, while relative decreases in miss rates were obtained through use of GTQs, the absolute levels of miss rates (38% and 28%, in C1a and C2a respectively) might be indicative of annotation items requiring further scrutiny and/or ease in task criteria definition. Our annotation task contained a combination of both demanding visual search and items with low ground truth base rates. Thus, it would be logical or even possibly more natural for a rater to adopt a conservative strategy when faced with annotation uncertainty (i.e., not checking a box unless they have explicitly seen something). Relatedly, the high miss rates may reflect a bias due to the fact that all items were by default unchecked (absent) requiring checking as needed, rather than being checked (present) requiring unchecking as needed. Indeed, complexities in universal instructions, clear coding rule descriptions, and controlled balancing of default absence/presence question valences could be a relevant concern in crowdsourcing annotations from large, diverse, and remote participant populations without local remediation of a real-time physically present experimenter. However, it should be noted that we did not use any CrowdFlower geography/language settings and thus kept this aspect equally random across our external worker conditions so as not to confound our relative evaluations regarding potential benefits of GTQs.

Secondly, the specific items of the coding scheme created and used in the present study may be challenged further than issues of clarity towards aspects of organization and inter-item independence. The item checkboxes within a Judgment were pre-tested and arranged by probable frequencies of occurrence such that categorization speeds might benefit from predictable and likely emergent patterns of responses. Thus, the repetitive and non-random ordering of items may be a source of bias towards consistency (although, again it should be noted that the same structure was presented to both GTQ and non-GTQ condition groups).

Lastly, several dependency relations existed between items which may degrade the power of some of the analyses of the present study. For example, several items pertained to the identification of object classes (cars, motorcycles, bicycles, and pedestrians, respectively) that upon selection, each expanded with sub-item location information (i.e., leading, oncoming, passing, crossing, parking). For cases where only one object from the class was present, the sub-item location information thus became mutually exclusive rather than independent. As another example, items pertaining to actions of other vehicles such as “Lane change by another vehicle” and “Turning on/off between this and any other road by another vehicle” logically depend on presence of another vehicle and thus retain relations to ratings of item vehicle class identification.

More traditional and established methods for interrater reliability (e.g. Cohen’s/Fleiss’ kappa) were not pursued. The reason for that is the difficulty of determining a chance agreement for our Judgments that contained a composite of yes/no decisions with inter-item dependencies as

described above. Instead, simpler measures of consistency, such as the phi coefficient and the proportion of unanimous Judgments, were used. Further studies with CrowdFlower more specific to questions of validity and reliability might limit such complexities in advance, sacrificing some annotation meaning in favor of stricter control, standard analyses, and afforded reflection regarding the broader annotation literature. Additionally, further assessments of the ground truth reliability of our internal rating team (beyond the single rater repetitions of the analysis in 3.2.1) would be desirable in future work. For now, the reliability agreements observed in our approach (Fig. 3.1.4) appear qualitatively consistent with levels from previous image annotation work (Nowak & Ruger, 2010; containing 53 annotations per image across a set of 99 without presuming the existence of two persons that annotated the whole set of images). Specifically, in comparison to the average identical accuracy they obtained of 0.906, following their Equation 2, we computed our own average unanimous annotation accuracies respectively as 0.941 (section 3.2.3, Fig. 3.1.4) and 0.906 (section 3.2.4).

Multiple ethical and privacy concerns can be raised in consideration of methods that employ crowdworkers with human annotation of naturalistic driving video data. Some of these may not be new and include attempting to anonymize video data in the sense that specific combinations of sensitive information are not presented in combination to result in personally identifiable information from both aspects of the drive (time, date, location, etc.) along with aspects of driver identity (name, face, home/work address, etc.). A major difference between the present method and the classical way of annotating naturalistic driving data is that in the present method the task is outsourced to crowdworkers who are themselves anonymous and residing in different countries, while in the classical way the annotation is done by trained team members who are typically local and known/approved by the principal investigator(s). Aside from the annotation integrity (accuracy/consistency) concerns specifically addressed in the experimental design and results of the present study, other new challenges are worth discussing such as legal requirements of the handling of data. In the present study, the video data were obtained from public sources, which is uncommon within traditional NDS approaches. Thus, any terms and conditions regarding data sharing, ownership, and viewership restrictions put in place a priori by the responsible parties would need to be considered and respected so as not to be violated. Additionally, the regulations and policies pertaining to the online reproduction/distribution of (video) data specific to each country or online hosting community should be adhered to, and this includes the presentation of potentially disturbing images such as might be the case with automobile crashes/accidents or illegal driving behavior.

A few positive privacy points regarding the present method are interesting to consider as well. Because the annotating work is distributed across many crowdworkers in distal locations, a relatively small amount of the total data is restrictively released to single/isolated persons at a time. For example, in the present study, only random 3-second clips from randomly different drives and randomly different drivers were distributed. Accordingly, it becomes much less likely that a crowdworker can come to recognize a driver's travel patterns or other aspects that may pose risks to privacy. This compares favorably in contrast to a classical annotation perspective where a single or smaller group of annotators may more likely become familiar with the travel patterns contained within the data. Additionally, the present study does not propose to share all data (e.g., geospecific, CANBUS, etc.) as may be accessible to classical annotators in naturalistic research but to selectively distribute only pieces of the full dataset (i.e., herein only video annotation was outsourced and only that of forward facing cameras from public roads where filming is allowed). Lastly, crowdworkers

themselves are employed under certain terms of service to which they must accept and abide (e.g., <https://www.crowdflower.com/legal/>). If crowdworkers were to violate such terms (e.g., share proprietary data) they would be subject to consequences not limited to but including the likes of losing their worker privileges such as payment, membership, etc.

An increasing amount of real-life driving videos are being recorded both within naturalistic driving studies as well as from public channels of user generated content. For example, at the start of conducting the current research, there were approximately 795,000 returns for the term “dashcam” on YouTube (November 19, 2015). Upon presenting this work at the international conference for Road Safety on Five Continents (May 19, 2016), there were 1.13 million returns for the same search (i.e., +42% increase in about half a year), and by the time of manuscript revisions (August 8, 2017), a total of 4.26 million were available (i.e., +436% increase in less than 2 years’ time). Categorizing such expansive data sets can be a costly and time-consuming manual process. One solution is to train automated algorithms to conduct coding tasks such as in machine learning and classification. However, such algorithms themselves often require some diligently pre-labeled examples for their own accuracy and only through diverse training sets may overcome common challenges of overfitting. Under the correct circumstances (e.g., open-access data) and quality control settings (i.e., the construction and use of GTQs), Crowdsourcing tools like CrowdFlower appear to have the potential for delivering equivalent accuracy and reliability utility as locally trained humans. It is therefore recommended that future driving research and ultimately driving safety itself might benefit from exploiting increasingly large scale and publically available data sets through embracing and channeling a growing global pool of human resources.

Acknowledgments

The research presented in this paper was conducted within the project HFAuto – Human Factors of Automated Driving (PITN-GA-2013-605817).

References

Bazilinsky, P., & De Winter, J. C. F. (2015). Auditory interfaces in automated driving: An international survey. *PeerJ Computer Science*. 1:e13.

Borowsky, A., Shinar, D., & Oron-Gilad, T. (2010). Age, skill, and hazard perception in driving. *Accident Analysis and Prevention*, 42, 1240-1249.

Chapman, P., Van Loon, E., Trawley, S., & Crundall, D. (2007). A comparison of drivers' eye movements in filmed and simulated dangerous driving situations. *Behavioral research in road safety, Seventeenth seminar*. London: Department for Transport.

Crundall, D., Underwood, G., & Chapman, P. (1999). Driving experience and the functional field of view. *Perception*, 28, 1075-1087.

Gosling, S., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology*, 66, 877-902.

Kyriakidis, M., Happee, R., & De Winter, J. (2015). Public opinion on automated driving: Results of an international questionnaire among 5,000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour*, 32, 127-140.

Le, J., Edmonds, A., Hester, V., & Biewald, L. (2010). Ensuring quality in crowdsourced search relevance evaluation: The effects of training question distribution. *Proceedings of the SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation (CSE 2010)*.

Nowak, S., & Ruger, S. (2010). How reliable are annotations via crowdsourcing: A study about inter annotator agreement for multi-label image annotation. *Proceedings of the International Conference on Multimedia Information Retrieval*, ACM, 557-566.

Regan, M., Williamson, A., Grzebieta, R., & Tao, L. (2012). Naturalistic driving studies: literature review and planning for the Australian naturalistic driving study. *Australasian College of Road Safety Conference*, Sydney, New South Wales, Australia.

Smucker, M., & Jethani, C. (2011). The Crowd vs. the Lab: A comparison of crowd-sourced and university laboratory participant behavior. *Proceedings of the SIGIR 2011 Workshop on Crowdsourcing for Information Retrieval (CIR 2011)*.

Soleymani, M., & Larson, M. (2010). Crowdsourcing for affective annotation of video: Development of a viewer-reported boredom corpus. In V. Carvalho, M. Lease, & E. Yilmaz (Eds.). *Proceedings of the SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation (CSE 2010)*.

University of Ottawa (2014). Validity and Reliability of Measurements: 1. Core knowledge. Retrieved from http://www.med.uottawa.ca/sim/data/Measurement_validity.htm

Appendix A. Coding Instructions/Training Material

Watching Very Short Videos And Marking Yes/no Checkboxes Confeds

Instructions ▾

Overview

Watch short driving video segments (3 seconds each) and select the elements that can be seen in the video segment. This should be a simple and easy objective job task and not much if any subjectivity, matter of opinion, or "thinking" involved at all.

We are trying to train a computer to recognize things in these videos and match like driving situations to each other but first we need to label what is actually in them so we are trying to use crowdsourced human eyes to objectively say what is there or not.

Each video categorization has been previously measured to take people about 1.5 minutes or less on average.

We provide

URL links to specific driving video segments ("Play Video" button)

Lists of true/false items (checkboxes)

Process/Procedure

Watch and review the video segment provided (3 seconds each) FULL SCREEN IS RECOMMENDED. Replay and pause video as often as you need. This is NOT a memory test. Check all that apply.

For the first half of the questions (1 through 7), we would like categorizations of what other vehicles/pedestrians are present within that driving video segment based on 5 different possible locations (directions of travel) relative to the driver whose vehicle the video was filmed from. Click on the check boxes to denote various vehicle/pedestrian presence by its location/position/direction-of-travel.

- a) in front of and in the same lane
- b) traveling in an oncoming/opposite direction
- c) traveling in the same direction whether alongside or ahead
- d) traveling at any different angles
- e) parked, parking, un-parking

For the second half of the questions (8a through 8k), we would like categorizations of other miscellaneous aspects within that same video segment. **Click on the checkboxes to indicate which elements are "Obviously Visible" within that video segment.** This may include things the vehicles of the videos do not actually reach or complete, but which you don't have to squint for under a magnifying glass.

Please note:

- "8b ... Just straight road (no angles/bends/curves in the entirety of the visible road of travel)" applies not only to the portion of the road driven but also that which is "Obviously Visible" anywhere within that video segment including the road ahead.
- Lane changes or turns don't have to be complete to count.
- Categorizations should span the full 3 seconds of the segment (watch out for things that are there in the first moments even if they shortly disappear due to motion of the video)

*****Disclaimer***:** Unfortunately some people cheat by clicking randomly or using computer programs to complete jobs. Please *****ALWAYS***** check the check box to confirm you are a diligent human contributor (it is located at the end of the "other vehicles/pedestrians" section). Also, please *****NEVER***** check any boxes that ask you to leave them empty, off, or unchecked (these occur at the beginning of each section and in the middle of the first section underneath group 3). Don't worry, these will be very very obvious! However, if too many are missed you will be excluded and not paid, so please no random clicking!

Steps

a) check all that apply ...

[Play Video](http://ttfa100.pandoraapps.net/showVideo.html?st=90&et=93&vid=5f6ykkcjruA&l=yes&lnf=1000000&ap=no)

Step 1
Click button to open URL in new tab/window

NEVER check this box; leave it unchecked/empty/off***

1) I can see one or more Cars/Trucks/Vans/Buses in this video segment...

2) I can see one or more Motorcycles/Scooters/Mopeds in this video segment...

3) ***NEVER check any checkboxes in group 3; leave this one and its sub parts all unchecked/empty/off***

4) I can see one or more Bicycles in this video segment ...

5) I can see one or more Pedestrians in this video segment ...

6) No one else present; this driver is alone

7) ***ALWAYS check this box to confirm you are a diligent human contributor***

8) Which elements are contained in THIS driving video segment?

8a... ***NEVER check this checkbox; leave it unchecked/empty/off***

8b... Just straight road (no bends/curves visible)

8c... More than one lane per direction of travel

8d... Any signs/symbols facing driver (road signs, billboards, traffic lights, building names, ads, etc.)

8e... Painted communication on any visible road surface (includes crosswalks, arrows, etc. but NOT lane boundary/edge line)

8f... Lane change by this driver

8g... Lane change by another vehicle

8h... Turning on/off (between this and any other road) by THIS driver

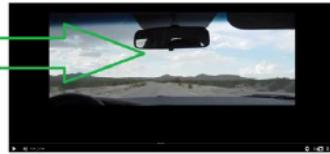
8i... Turning on/off between this and any other road by another vehicle

8j... This driver slowing to a stop, being stopped, or pulling away from a stop

8k... Editing alterations in the video file (discontinuity, added text, pauses, slow motion, sped up sections, etc.)

8l... None of these misc. elements are present in this video segment

Step 2
Watch/repeat 3 second long driving video segment



Step 3
Check all that apply regarding the obviously visible aspects of this driving video segment

- Don't check boxes that say "NEVER" check ...

- Be sure to check the common box that says "ALWAYS check ..."

b) check all that apply, including sub statements as they appear...

Step 1
Click button to open URL in new tab/window

Step 2
Watch/repeat 3 second long driving video segment

Step 3
Check all that apply regarding the obviously visible aspects of this driving video segment

Checklist items (with annotations from the image):

- **NEVER check this checkbox: leave it unchecked/empty/off**
- 1) I can see one or more Cars/Trucks/Vans/Bus in this video segment ...
- 2) I can see one or more Motorcycles/Scooters/Scrapeds in this video segment ...
- 3) ***NEVER check any checkboxes in group 3: leave this one and its sub-parts all unchecked/empty/off***
- 4) I can see one or more Bicycles in this video segment ...
- 5) I can see one or more Pedestrians in this video segment ...
- 6) No one else present, this driver is alone.
- 7) ***ALWAYS check this box to confirm you are a diligent human contributor***
- 8) Which elements are contained in THIS driving video segment?**
 - 8a ... ***NEVER check this box: leave it unchecked/empty/off***
 - 8b ... Just straight road (no bends/curves visible)
 - 8c ... More than one lane per direction of travel
 - 8d ... Any signs/signals facing driver (road signs, billboards, traffic lights, building names, ads, etc.)
 - 8e ... Pedestrians/cyclists on any visible road surface (includes crosswalks, arrows, etc. but NOT line boundary/edge line)
 - 8f ... Lane change by this driver
 - 8g ... Lane change by another vehicle
 - 8h ... Turning on/off between this and any other road by THIS driver
 - 8i ... Turning on/off between this and any other road by another vehicle
 - 8j ... This driver slowing to a stop, being stopped, or pulling away from a stop
 - 8k ... Editing alterations in the video file (discontinuity, added text, pauses, slow motion, sped up sections, etc.)
 - 8l ... None of these miles elements are present in this video segment

Reasoning for Step 3 (from the image):

1) 1a ... they are leading (same lane ahead)
 1b ... they are oncoming (traveling opposite direction)
 1c ... they are passing or being passed (traveling same direction)
 1d ... they are turning (traveling on a road that intersects)
 1e ... they are parked (parking, or un-parking/pulling out)

5) 5a ... they are leading (same lane ahead)
 5b ... they are oncoming (traveling opposite direction)
 5c ... they are passing or being passed (traveling same direction)
 5d ... they are crossing (traveling on a road that intersects)
 5e ... they are parked (parking, or un-parking/pulling out)

Tips

The URL link may appear broken across multiple lines, so be sure to copy/paste the entire URL link all the way from "http://ttfa..." and ending with "...&" if needed

Use "full screen" to see the video in a larger view.

If you miss any in QUIZ mode, please read the provided answers and reasoning carefully as it should help clarify what we mean/expect by our various categorization items. Also refer to Extended Sample Set below as needed for visual examples of various items.

Thank You!

Your help on this task is greatly appreciated!

Extended Sample Set (reference as needed)

Note: more than one item may apply within the same video segment (always check **all** that apply). In these examples, we have used highlights only to show some of the possible clues or cues that should help you know to mark the current specific check box item as true.

1a) Car/Van/Truck/Bus ... they are traveling in the same direction in the same lane ahead (leading)



1b) Car/Van/Truck/Bus ... they are traveling in the opposite direction (oncoming)





1c) Car/Van/Truck/Bus ... they are traveling in the same direction (passing, being passed, pass-able)



1d) Car/Van/Truck/Bus ... they are traveling on a road that intersects (crossing)



1e) Car/Van/Truck/Bus ... they are parked, parking, pulling out/unparking



2e) Motorcycles/Scooters/Mopeds ... they are parked, parking, pulling out/unparking



4c) Bicycles ... they are traveling in the same direction (passing, being passed, pass-able)



5b) Pedestrians ... they are traveling in the opposite direction (oncoming)



5c) Pedestrians ... they are traveling in the same direction (passing, being passed, pass-able)



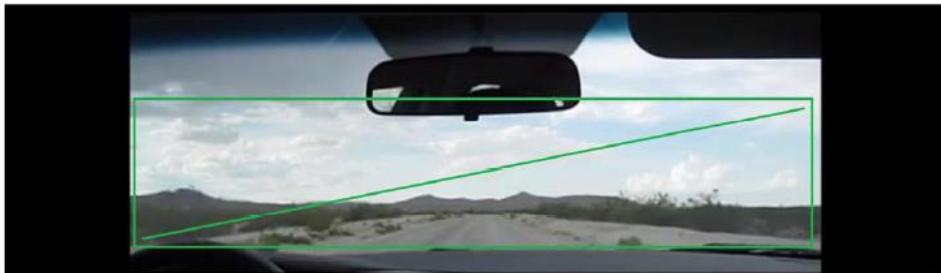
5d) Pedestrians ... they are traveling on a road that intersects (crossing)



5e) Pedestrians ... they are standing still or otherwise stationary



6) None of the above. No external vehicle/pedestrian present. This driver is alone.



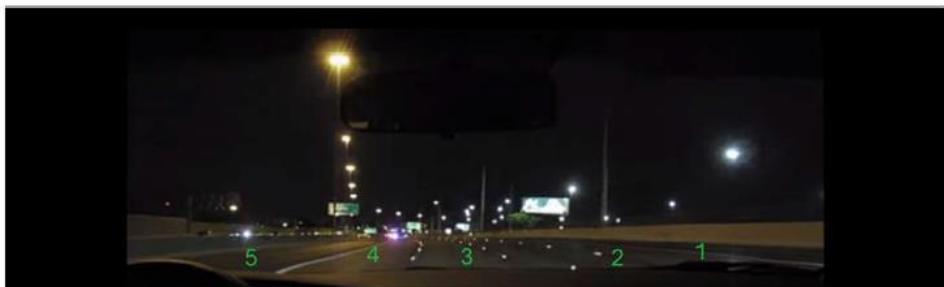
8B) ... Just straight road (no angles/bends/curves in the entirety of the visible road of travel)



8C) ... More than one lane per either direction of travel

Please note that some lane division markings may vary (white/yellow) between different roads. However, please mark 8c only if there is little to no ambiguity in the multi lane situation. For example, immediately above in the images of "8b" at different times with other vehicles present or not, or even with and without a visible dividing lane line at all, it might be hard to tell if it should be considered either a single lane of travel in both directions (8c would not apply = false) or two lanes in the direction of travel of the driver (8c would apply = true). Please assume that the later second to be a very RARE CASE and you might only expect it when there is another parallel roadway for the other direction of

travel (i.e. a "divided highway" situation). Furthermore if the road has no lane division markings then it has no "lanes" and so 8c = false. Please, reserve the marking of 8c = true for clearly obvious and unambiguous multiple lanes per either direction of travel (see below with lanes counted out in green numbers ascending from the edge to the center of the road separately per direction of travel).

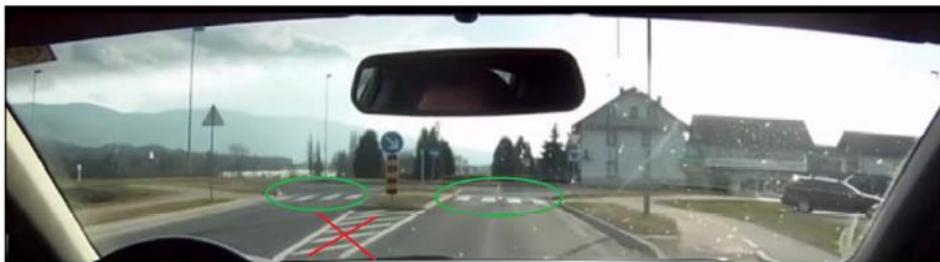


8D) ... Any signs/signals facing driver (road signs, billboards, traffic lights, building names, ads, etc.)

Note: if you can make out colors, text, symbols, pictures, etc. on the sign/signal then for our purposes here it is "facing the driver" even if it is not 100% legible; the only ones you don't count are those that are just the backs of signs (e.g. facing the opposite direction, oncoming traffic).



8E) ... Painted communication on any visible road surface (includes crosswalks, arrows, etc. but NOT lane boundary/edge info)





8F) ... Lane change by this driver



8G) ... Lane change by another vehicle



8H) ... Turning on/off between this and any other road by THIS driver





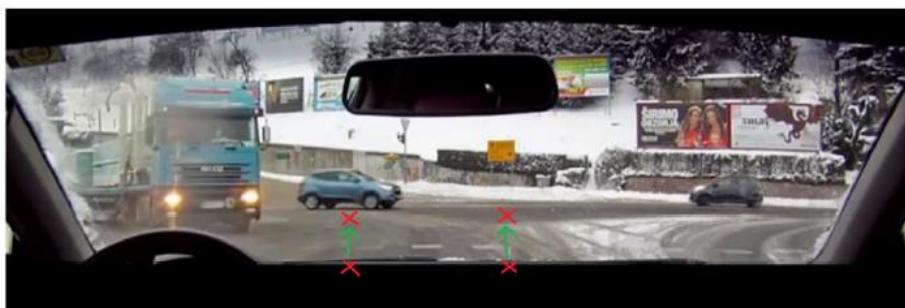
8I) ... Turning on/off between this and any other road by another vehicle



8J) ... This driver is slowing to a stop, is stopped, or pulling away from a stop.

Note: The red crosses here are meant as possible places you might notice deceleration or other motion patterns indicative of this item. The green circles are other possible/probable

clues of stopping contexts in common driving situations.



8K) ... Editing alterations in the video file (discontinuity, added text, pauses, slow motion, sped up sections, etc.)



8L) ... None of these miscellaneous elements are present in this video segment



Thank you.

[http://ttfaloopandrepeat.appspot.com/showVideo.html?
st=90&et=93&vId=5HiykkcjruA&l=yes&lnf=1000000&ap=no](http://ttfaloopandrepeat.appspot.com/showVideo.html?st=90&et=93&vId=5HiykkcjruA&l=yes&lnf=1000000&ap=no)

Play Video ([http://ttfaloopandrepeat.appspot.com/showVideo.html?
st=90&et=93&vId=5HiykkcjruA&l=yes&lnf=1000000&ap=no](http://ttfaloopandrepeat.appspot.com/showVideo.html?st=90&et=93&vId=5HiykkcjruA&l=yes&lnf=1000000&ap=no))

NEVER check this checkbox: leave it unchecked/empty/off

1) I can see one or more Cars/Trucks/Vans/Buses in this video segment ...

1)

- 1a ... they are traveling in the same direction in the same lane ahead (leading)
- 1b ... they are traveling in the opposite direction (oncoming)
- 1c ... they are traveling in the same direction (passing, being passed, pass-able)
- 1d ... they are traveling on a road that intersects (crossing)
- 1e ... they are parked (parking, or un-parking/pulling out)

2) I can see one or more Motorcycles/Scooters/Mopeds in this video segment ...

2)

- 2a ... they are traveling in the same direction in the same lane ahead (leading)
- 2b ... they are traveling in the opposite direction (oncoming)
- 2c ... they are traveling in the same direction (passing, being passed, pass-able)
- 2d ... they are traveling on a road that intersects (crossing)
- 2e ... they are parked (parking, or un-parking/pulling out)

3) ***NEVER check any checkboxes in group 3: leave this one and its sub parts all

unchecked/empty/off***

- 4) I can see one or more Bicycles in this video segment ...
- 5) I can see one or more Pedestrians in this video segment ...
- 6) None of the above. No external vehicle/pedestrian present. This driver is alone
- 7) ***ALWAYS check this box to confirm you are a diligent human contributor***

8) Which elements are contained in THIS driving video segment?

- 8a ... ***NEVER check this checkbox: leave it unchecked/empty/off***
- 8b ... Just straight road (no bends/curves in the entirety of the visible road of travel)
- 8c ... More than one lane per either direction of travel
- 8d ... Any signs/signals facing driver (road signs, billboards, traffic lights, building names, ads, etc.)
- 8e ... Painted communication on any visible road surface (includes crosswalks, arrows, etc. but NOT lane boundary/edge info)
- 8f ... Lane change by this driver
- 8g ... Lane change by another vehicle
- 8h ... Turning on/off between this and any other road by THIS driver
- 8i ... Turning on/off between this and any other road by another vehicle
- 8j ... This driver slowing to a stop, is stopped, or pulling away from a stop
- 8k ... Editing alterations in the video file (discontinuity, added text, pauses, slow motion, sped up sections, etc.)
- 8l ... None of these miscellaneous elements are present in this video segment

Comments?

[http://ttfaloopandrepeat.appspot.com/showVideo.html?
st=99&et=102&vId=Ge0a27WR6WI&l=yes&lInf=1000000&ap=no](http://ttfaloopandrepeat.appspot.com/showVideo.html?st=99&et=102&vId=Ge0a27WR6WI&l=yes&lInf=1000000&ap=no)

Play Video ([http://ttfaloopandrepeat.appspot.com/showVideo.html?
st=99&et=102&vId=Ge0a27WR6WI&l=yes&lInf=1000000&ap=no](http://ttfaloopandrepeat.appspot.com/showVideo.html?st=99&et=102&vId=Ge0a27WR6WI&l=yes&lInf=1000000&ap=no))

- ***NEVER check this checkbox: leave it unchecked/empty/off***

Appendix B. Developed Driving Research Tools

B.1. Driving scene content annotation

In order to supply certain categories of driving situations for future research experiments (e.g., Chapter 3.2), and to validate the accuracy and reliability of driving situation categorizations from online crowdsourced workers (i.e., Chapter 3.1), it was necessary to devise and employ a driving scene content coding scheme (Figure 3.1.B.1). Goals of the coding scheme were that it be fairly comprehensive in regards to potentially interesting driving scene factors on a prototypical level, that the items would be objectively understandable to identify, and that items would be arranged in a manner such as to facilitate fast annotations with minimal effort. Two major groups of probabilistically ordered binary checkboxes were implemented. The first major group was in regards to various kinds of road-users (with expandable position/direction of travel details) while the second major group pertained to more miscellaneous infrastructural or behavioral aspects. Resulting average annotation durations were approximately 75 seconds each for a 3-second long driving video clip.

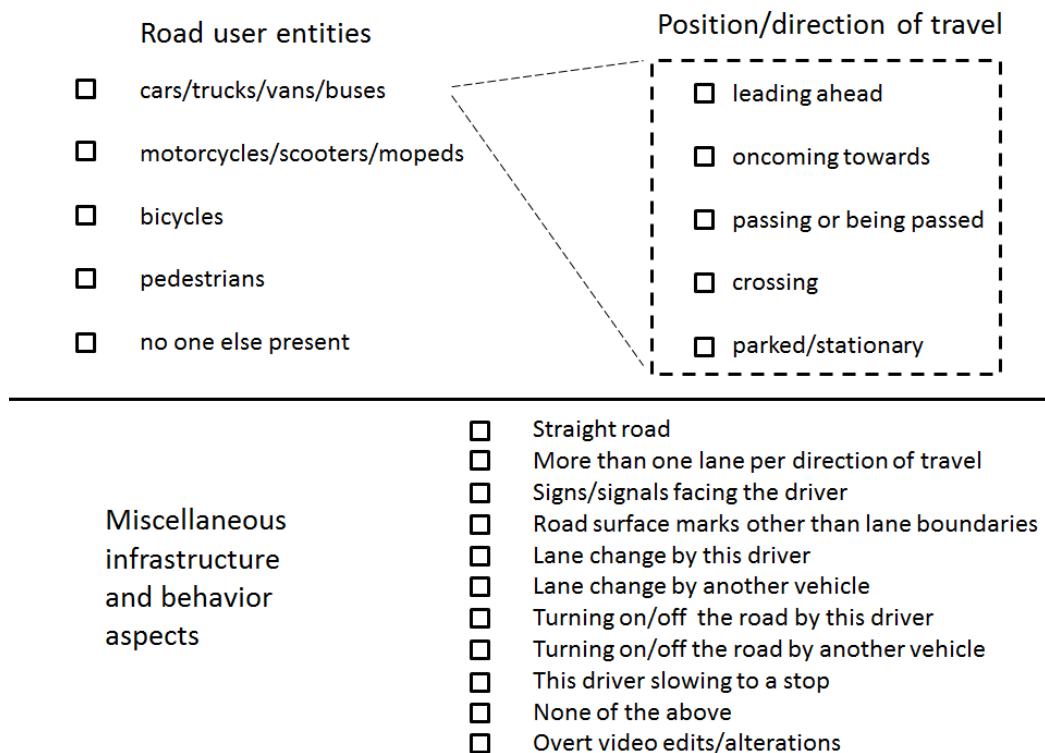


Figure 3.1.B.1. Driving scene content categorization coding scheme of binary checkboxes. Item arrangement included a first major grouping of road user entities that included an automatically expanding set of items for detailing their respective positions/directions of travel, and a second major grouping of miscellaneous infrastructure and behavior aspects.

B.2. Driving video clip selector

Ratings from the experiment of 3.1 generated 12,766 driving video segments with content annotations. Rather than storing each of these 3-second long video clips individually, a GUI (Figure 3.1.B.2) was devised and implemented from which to automatically parse (play/save) segments from any of the original 49 source videos in accordance to specified items contained in a hard-coded library of annotations. The standalone executable and source code files have been made freely available within the online repository of Zenodo at:

(1) <http://doi.org/10.5281/zenodo.2542314> (github software release)

(2) <http://doi.org/10.5281/zenodo.2542275> (repository for the 49 source videos)

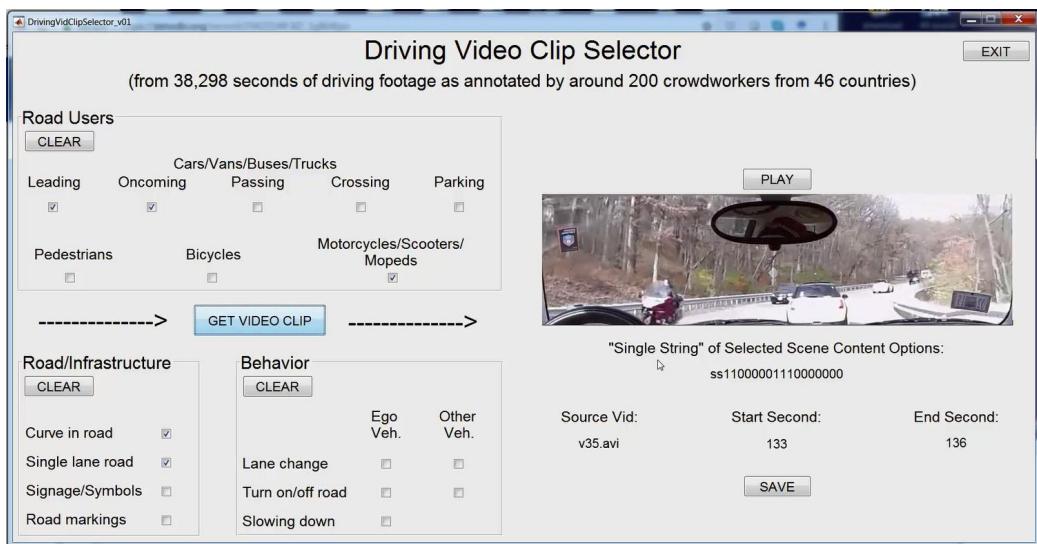
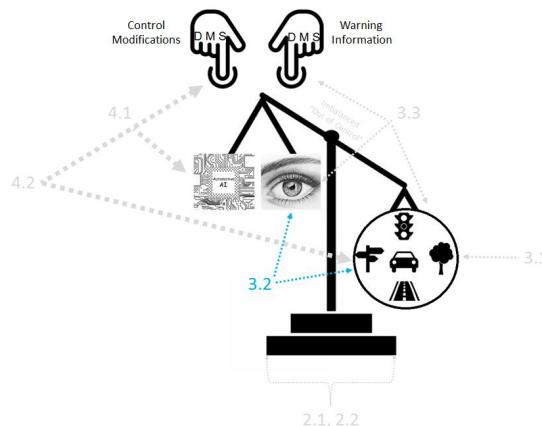


Figure 3.1.B.2. A standalone GUI for retrieving driving scenes of specified contents (i.e., here as annotated as 3-second duration video clips within the experiment of 3.1 from a driving video library of 49 separate videos).

Chap. 3.2) Estimating Driver Readiness from Situated Eye Movements: Prediction of workload and attention requirements from quantification of driving scene components



In regards to the overall thesis big picture, this experiment serves to relate driver perceived workload estimates and common eye movement measures with specific quantifiable visual properties of various driving scenes. The corpus of video annotations from Chap 3.1 supplied a range of driving scene contents/demands to be used as stimuli for Chap 3.2. Compared to the on-road Chap 3.3 study, conditions/measurements could be manipulated with a higher level of precision and control. Results showed road angle curvature to consistently be the strongest predictor of workload and eye movements, and amount of other road users likely to be of next greatest importance (when compared to other visible driving scene aspects like signage/symbols, buildings, etc). Saccadic amplitude was found to be the most sensitive eye measure (in comparison to fixation duration and pupil size) for representing workload demands of driving scenes. Such relational knowledge supplies predictive regression models and data to support fitness-to-drive driver monitoring systems. Thus, assessments are enabled towards determining if a person's eyes are moving appropriately *enough* provided the measurable contents (visual demands) of a specific scene they are driving within or when about to receive driving control from an automated/autonomous driving system (e.g., regardless of which human or computer agent initiates the request for the transition of control).

Adapted from:

Cabral, C.D.D., Happee, R., & de Winter, J.C.F. (under review). Estimating driver readiness from situated eye movements: Prediction of workload and attention requirements from quantification of driving scene components.

Abstract

The designs of higher levels of driving automation include times where drivers will uptake (upon request or voluntarily) some increased amount of driving responsibility while the vehicle is in motion. Thus, a key research and application development question concerns quantifying what impact contents of different driving task environments might be expected to have in establishing a driver's readiness to drive. This paper investigated predictability of nominal driver workload and attention while viewing and rating driving scenes with differing amounts of visible scene components: road curvature, road surface area, road users, signs/symbols, buildings/infrastructure, and vegetation/trees. We presented 60 randomly ordered dash cam video clips (3 s duration) and recorded the eyes of 15 participants who were tasked to provide ratings between 0 and 100 to the question of "how much effort for you to take control and drive within that segment?". Multiple linear regression models were derived and found to significantly improve prediction of workload ratings and eye movements from differently weighted combinations of the visible scene component factors. Road angle curvature was consistently the strongest predictor of workload and eye movements, and amount of other road users appeared to be of second greatest importance. From workload and driving scene components, the highest amount of explainable variance in eye measures was found in saccadic amplitude as compared to fixation duration and pupil size. In conclusion, the present regression equations establish quantifiable relations between how much workload and attention different driving scenes might require. In future driver monitoring systems, such knowledge can help inform road-facing and driver-facing cameras to jointly establish and verify the adequacy of a driver's level of engagement.

1. Introduction

1.1 Background motivation

With the advent of driving automation systems (SAE, 2018), new human factors road safety challenges exist for assessing driver states due to their altered roles and responsibilities. With SAE Level 3 'Conditional Driving Automation' and SAE Level 4 'High Driving Automation' drivers will be removed from sustained involvement in the driving task until either being called back in, or at a point of voluntary uptake. As seen in Fig. 3.2.1 (from Petermeijer et al., 2016), a so-called 'Take Over' process involves a transitional phase, where attentional shifts and cognitive processing are expected to occur over a period of time, prior to increased conventional driver control activities. Conceptually, there can exist a buffer between eyes on road and the start of manual driving.

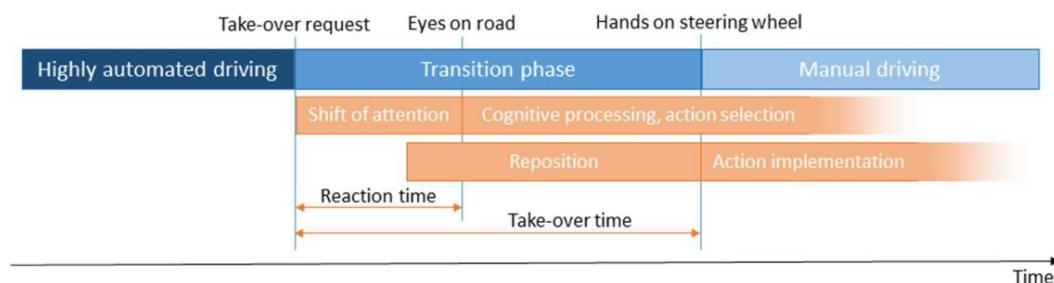


Figure 3.2.1. The take-over process from highly automated to manual driving, adopted from Petermeijer et al. (2016).

From a safety system assurance perspective, it is reasonable to expect that a Driving Monitoring System (DMS) layer could provide oversight during such a transition (e.g., assumptions of requests ought to be tested). Such a verification could transpire whether the direction of take over request transpires from Automation-Initiation towards Driver-Control or from Driver-Initiation towards Driver-Control (i.e., AIDC and DIDC respectively from Lu et al., 2016). For comprehensive consideration, it is important to note that neither of such AIDC/DIDC transitions necessarily connotes an emergency situation but might each be further classifiable as either nominal or critical with subsequent corresponding differences in aspects of typical task timing, events, and environmental characteristics.

A recently emerging body of literature has focused on establishing the timing requirements of transitions of control from automated driving systems to human drivers. Standards regarding AIDC transitions of control (aka. take over requests, dynamic driving task fallback, requests to intervene, etc.) have suggested that the human should be allotted some phase of fair lead-in time: '*with notice*' (NHTSA, 2017), '*sufficiently comfortable transition time*' (NHTSA, 2013), '*with a certain time buffer*' (BAST, 2012), '*At level 3, an ADS is capable of continuing to perform the DDT for at least several seconds after providing the fallback-ready user with a request to intervene*' (SAE, 2018). By reviewing automated to manual driving transition timings across 25 papers, Eriksson and Stanton (2017) determined that an average allotment period (until a critical event) was around 6 seconds and that an average reaction time (to take back vehicular control) was around 3 seconds. Their own empirical measurements with non-critical transitions (with and without secondary tasks) found substantially increased timing requirements up to 25.75 seconds to resume control from automated driving in normal conditions. In their discussion, Eriksson and Stanton (2017) recommend a case for adaptive automation that modulates a take over request lead time by

detection of driver gaze, such that for example, a few additional seconds might be provided to a driver ahead of resuming control.

Obtainment of situation awareness (cf. Endsley, 1995) is intuitively presumed as a requisite target threshold for establishing driver readiness, however specifics of the cognitive constituents of that construct can be problematic and might be beyond what is necessary for some initial practical application benefits. Mok et al. (2015) argue that accidents may result if drivers do not sufficiently assess the situation prior to taking control. Lu et al. (2016) propose that the demands on the timing of the automated driving technology transitions are set by how much time drivers need for gaining situation awareness. For example, from viewing and then reconstructing portions of simulated driving scenes (i.e., after periods of inattentiveness), situation awareness for positions of other vehicles reached saturation between 7 and 12 seconds and for their velocities at a range beyond 20 seconds (Lu et al., 2016). Such assumptions and results, however, can raise questions of how good is good enough when it comes to defining a complete mental grip and/or what parts of the situation are relevantly necessary for establishing adequate levels of awareness (cf. the MiRA theory in Kircher & Ahlstrom, 2016). A standard situation awareness measurement is the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1988) with a body of literature evidencing positive association with performance (Salmon et al., 2009; Gardner et al., 2017; Prince et al., 2007; Gugerty, 1997; McGowan & Banbury, 2004; Loft et al., 2015; O'Brien & O'Hare, 2007). However, the practical utility of the SAGAT has also seen contentious results concerning its predictive validity with performance (Durso et al., 2006; Durso et al., 1998; Pierce et al., 2008; Strybel et al., 2008; Cummings and Guerlain, 2007; Ikuma et al., 2014), and has been criticized for its reliance on memory (Gutzwiller et al., 2013) and on explicit representations amongst other limitations (Stanton et al., 2015; De Winter et al., in press).

An earlier emotive impression before conscious expression might be more accessible/practical as a cognitive construct that would be useful in modeling driver readiness assessments for uptake of driving control. Stanton and Young (2000) developed and proposed a psychological model of driving automation in which situation awareness is the last in a chain of cognitive constructs and with a mental workload construct feeding into it. More recently, Heikoop et al. (2015) updated that model from a systematic literature search of driving automation papers and a subsequent quantification of reported links between psychological constructs. Notably, the updated Heikoop et al. (2015) model maintains the same directional relation of a later positioned situation awareness that is fed from an earlier positioned mental workload (with a newly interceding construct of attention). A widely adopted standardized measurement of mental workload is the NASA TLX (Hart & Staveland, 1988) and in essence consists of high/low scales for subjectively rating demands. In contrast, the situation awareness of the SAGAT (as previously introduced above), consists of presumably later or higher levels of conscious representation and recall. Moreover, malleable attention resources theory (MART) from Young and Stanton (2002) has posited a nominal human ability to muster/diminish attentional pools in an adaptive manner thereby shaping information processing capacities to match and meet present demands. Thus, in modelling a transition from automated to manual driving control, presumed driving effort appears to be a reasonable starting place to parameterize as a construct from which attention and situation awareness would be expected to follow.

1.2 Driver eyes, workload, and scene relations

Measurements of the eyes of drivers might reasonably contain important precursor information towards assessing a driver's readiness to drive. The information that drivers use is predominantly considered to be visual and could benefit from enhanced quantitative frameworks (Sivak, 1996). In a review of 50 years of driving safety research, Lee (2008) concluded that most accidents occur because '*drivers fail to look at the right thing at the right time*'. Senders et al. (1967) empirically investigated the amount of visual attention requested from human drivers for different roadways via an inverse occlusion technique to develop values of some of the parameters of a mathematical model of attentional demand. Subsequent driver visual workload studies have continued to employ such occlusion techniques to scientifically evaluate driving safety requirements (Van der Horst, 2004) and persist to present day methodological studies where they are advocated for use in combination with think aloud protocols and eye tracking (Kicher & Ahlstrom, 2018). According to a strong form of an eye-mind hypothesis, gaze direction is a perfect correlate of cognitive activity (Just & Carpenter, 1980). According to Moray (1990, 1993), information acquisition while driving is limited by eye movement characteristics, and attentional changes in dynamic real environments are equivalent, in operative terms, to changes in eye fixations. Thus, it should be theoretically possible to identify cognitive processes from eye movements if the environment is known and the task constrained.

Eye tracking parameters are often used as correlates of mental workload (Ries et al. 2018) and have been reviewed specifically for the case of drivers (Marquart et al., 2015). In contrast to mixed results regarding blink rates (Table 3.2.1), the eye measures of pupil size, fixation duration, and saccade amplitude show an apparent consensus of directional consistency in driving studies regarding mental workload and thus are introduced in turn below and taken as candidates for the present modeling purposes.

Table 3.2.1. Relation of eye-related physiological measures and drivers' mental workload (adapted from Marquart et al., 2015)

Measure	Mental workload (+)
Blinks	+ / -
Pupils	Dilation
	ICA
Fixations	+
Saccades	-

Note. ICA = Index of Cognitive Activity (Marshall 2000, 20002).

1.2.1. Pupil Size

In a driving simulator study, Palinko et al. (2010) found increased pupil dilation diameters to be positively associated with increased driver cognitive loads and decreased driver performance. Hence, Palinko et al. (2010) concluded that pupillometry shows promise as a measure for changes in driver cognitive load in line with the seminal phenomenon of a task evoked pupillary response (TEPR) of Beatty (1982) where pupils dilate when people are faced with challenging cognitive tasks. For example, Ahern (1978) found increased pupil diameters with increasingly difficult mental multiplication problems and these results have been recently replicated with more modern day equipment by Marquart & De Winter (2015). In a simulated driving task, Schwalm et al. (2008) has applied an index of cognitive activity (ICA) that is based on changes in pupil dilation from Marshall

(2000, 2002) and found that ICA increases in situations with higher mental demand on the driver i.e., when performing a lane change maneuver or an additional secondary task.

1.2.2. Fixation Duration

In actual on-road driving, the complexity of the traffic environment produced a latency in eye movement suggesting a deeper processing at each fixation point (Miura, 1990) and Recarte & Nunes (2000) found longer fixations during spatial-imagery tasks. Older drivers (e.g., who presumably suffer degraded information processing capabilities and thus greater cognitive effort) compared to younger drivers, were found to exhibit increased fixation durations while viewing safety related areas of interest in pictures of real-life traffic scenes (Maltz & Shinar, 1999). Underwood et al. (2011) reviewed a series of driver hazard perception studies and found an increase in fixation durations akin to a weapon/threat focus of Loftus et al. (1987). In between a fixation and a saccade exists an eye measure known as a saccadic intrusion where the eye makes small shifts away and then back to the original fixation in a fast and jerky manner, thus an increase in saccadic intrusions is consistent with an increase in fixation durations and a decrease in saccadic amplitudes. Tokuda et al. (2011) measured saccadic intrusions where participants were instructed to examine pictures of a highway driving scene while completing a cognitively loading N-back task: higher mental workload was found to produce an a greater number of saccadic intrusions.

1.2.3. Saccade Amplitude

In controlled laboratory investigations of free viewing (not requiring any specific fixation or tracking), increased cognitive demands (i.e., of an auditory tone counting task) consistently decreased saccadic amplitude ranges across four separate experiments (May et al., 1990). In a driving simulator study, Tsai et al. (2007) found evidence of reduced ranges of scanning when subjects were dually tasked with driving and an auditory addition task. Increased mental demands for on-road drivers in instrumented vehicles were found to produce spatial gaze constriction via decreased gaze variances (Recarte & Nunes, 2000, 2003; Reimer, 2009). Using both an on-road instrumented vehicle and a driving simulator, Victor et al. (2005), found a decrease in standard deviation of gaze angle/position in the presence of an additional auditory task and on roadways of increased driving task complexity (curved over straight sections, rural over motorway roads, and on-road over simulator). Furthermore, Victor et al. (2005) found that higher visual demands (i.e., of an in-vehicle secondary task) increased gaze variance (i.e., away from the road).

In the updated psychological model of driving automation from Heikoop et al. (2015), task demands were found to positively affect mental workload and published driving theory presents both challenges and potential for predictive applications. It is reasonable to expect that some driving scenes might nominally be more or less difficult than others, even beneath/before consideration of additional complicating cases of emergency or safety critical scenarios. Road safety research may often include driving conditions of varyingly low/high complexity, but quantifications of their differences are more easily eschewed for intuitive qualitative characterization. In lessons learned from developing driving research scenes and scenarios, Papelis et al. (2003) argue that *'Often times, specifications about the characteristics of the ambient traffic or ambient environment are missing or incomplete. ... it is often the case that variations in these ambient characteristics of a scenario can make a drastic difference on how participants perceive the scenario.'* When it comes to detailed and specific accounts of driving task demands, examples from the theoretical literature may both problematically appear either considerably vague: *'road surface conditions, road infrastructure*

layout, visibility and the behaviour of other road users', 'normal conditions (e.g., daylight, dry road surface, sparse traffic, wide lane)' – Engstrom et al. (2013) or overly anecdotal 'when a driver sees two cars approaching on a two-lane road and the rear car swings out to pass with the intention of cutting in before the various paths meet' – Gibson and Crooks (1938). However, Michon (1979, 1985) conveyed and then elaborated upon a model of driving that has become widely adopted where complexity involves aspects of strategy - issues of overall trip planning/goals and risk acceptance; of tactical maneuvers - management of risk probabilities through negotiations like speeding up, slowing down, turning, and overtaking; and of operational control - the basic skills of steering and braking for lateral and longitudinal positioning. Under Michon's driving complexity categories, visible scene components might be categorized accordingly and expected to impact perceived driver mental workload: buildings/destinations and/or nature-scenic routes as strategic aspects; signage/symbology that govern rights of way priority and predict traffic behaviour interactions as tactical maneuver aspects; and lateral course and longitudinal collision conflicts (road curvature and traffic, respectively) as operational control aspects.

Previous empirical investigations show promise for ascertaining effects of driving scenes on drivers' mental workload. In the dissertation of De Waard (1996), the road environment and traffic demands were identified as complexity factors contributing to driver workload and trends were evidenced between baseline and loaded conditions in the predicted direction but did not obtain statistical significance. The road conditions included for environmental complexity considerations of De Waard (1996) were sections with and without motorway entrance and exits, sections with and without adjacent noise barrier walls, and rural roads through forests or open moorland. Steyvers et al. (1994) argued that driving is a well-defined task which has to be executed in an environment that is readily describable and has a clearly identifiable task context and thus approached their study of driving behaviour as an activity which could benefit from a computational approach. Participants were shown recordings (about 80s each) filmed from behind the windshield of a moving car, and with instructions to presume they were driving the car from which the film was recorded, were tasked with providing evaluations of experience. Driving films (of two lighting and two traffic conditions) were selected from two different roads with previously recorded lower/higher incidence of accidents: a more visually simplistic polder road (flat horizon with open fields of uniform vegetation size and density) and another rural road but with more varied visuals (contoured horizon of groups of bushes and forests with other segmented vegetation). From a reduction of qualitative attributes into subjective factor labels (i.e., '*hedonic value*', '*activational value*', and '*perceptual variation*') Steyvers et al. (1994) concluded that a combination of conditions and experiences accounted for previously unexplained single vehicle accidents on different kinds of roads.

For determining a direct relationship between driving scenes and the eyes of drivers, it is important to recognize that a competition or compromise of exogenous (bottom-up) and endogenous (top-down) factors has been a common and longstanding topic for the psychology of perception in general. In a treatise on active vs. passive visual search Tsotsos (1992) relayed that while a concept of active perception might have been relatively new to computer vision at the time, Helmholtz (1910) believed in perceptual hypotheses, the derivation of best interpretations given evidence, and in attentional mechanisms that guide processing even without eye movements. Some of the earliest evidence was obtained from Buswell (1935), where participants were asked to look at different types of artwork and as a result fixation positions were found to be highly regular and related to information in the pictures (e.g., preferences for people rather than backgrounds) and

thus by extension to perceptual and cognitive processing of the scene. Classic and widely cited scene viewing eye tracking work of Yarbus (1967) evidenced diverse scan patterns in the presence of different goal-directed instructions (e.g., give the age of a person, estimate their wealth, remember the positions and details of everything in the picture, etc.). Yarbus (1967) found similar but non identical eye movement patterns in free-viewing without any instructions and saw preference for particular areas/items of a scene (e.g., a face) and/or sub-parts of a face (eyes, nose, and mouth). Underwood and Radach (1998) have concluded that '*eye guidance appears as low-level as needed and as cognitive as possible for a given set of circumstances*'. Henderson & Hollingworth (1998) provide a comprehensive overview of literature surrounding eye movements during scene viewing and conclude with their saliency map framework wherein initial movements of the eyes are controlled by stimulus rather than cognitive features and after which saliency weights are modified to reflect relative cognitive interest of those regions (e.g., needs of perceptual and cognitive analysis of a region). A continuing debate about relative contributions of low- and high-level factors in targeting eye movements during scene viewing is given by Tatler (2009). Furthermore, Tatler et al. (2011) observed that the dominant framework regarding gaze allocation in scene viewing has been of image salience but that based on new principles of selection, frameworks of reward maximization and uncertainty reduction are also emerging. Blended visual sampling accounts can be found across highly cited work of Senders (1964, 1983), the Saliency-Effort-Expectancy-Value (SEEV) model introduced by Wickens et al. (2003), and a recent replication-extension study of Eisma et al. (2018).

1.3 Automated Driver Readiness Assessment Framework

Recently developed driving automation systems entail transitions of control back to human drivers in the middle of driving and thus present novel challenges of assessing driver readiness as introduced in the background motivation. From the reviewed literature above, the eyes of drivers appear to be reliably influenced from both cognitive aspects such as mental workload as well as from driving scene components. Recent reviews of empirical investigations for advanced driving automation (c.f. Ohn-bar & Trivedi, 2016) indicate trends towards automated vehicles utilizing cameras that point both outward at the driving scene as well as inward towards a vehicle's occupants. These vehicle cameras will likely come equipped with various increasingly available computer/machine vision capabilities, e.g., MathWorks (2018) and Krzywinski (2018), that through recent advances in the layered disciplines artificial intelligence, machine learning, and deep learning can be applied to automatically segment driving scenes e.g., Cityscapes Dataset (2018) as well as the interiors of vehicles, e.g., Eyeris (2018) and eyeSight (2018), which include tracking of the body, head, face, and eyes, etc. of different vehicle occupants, esp. for example, that of a current or would-be driver.

Thus, an adaptive control framework (Fig. 3.2.2) appears plausible and is proposed as a timely solution inspired by Wickens & Hollands (2000) for the present problem of estimating fitness-to-drive to modulate the transitional phases preceding a return of control to a human driver (cf. Fig. 3.2.1). Different measures may be taken from would-be drivers in real-time and the eyes are commonly presumed as a reliable indicator of attention under nominal circumstances. Such online measures can be compared against target reference eye measure predictions collected (again under nominal circumstances) as from environmental driving scene components directly (observed) and/or in conjunction with mental workload effort ratings (indirectly presumed). Mental workload can be assumed to be an earlier emotive psychological state adjacent to meeting driving task demands than fully conscious situation awareness, and hence a better choice for establishing

reference levels of situated fitted-ness. However, even if relatively more immediately accessible than situation awareness, mental workload (just as with any psychological construct) must ultimately be indirectly mediated, and thus potentially available only in partial manner, and/or influenced by other cognitive drivers (e.g., additional cognitive states and/or secondary tasks, etc.). Fitting into the lower right corner of the broader framework (and highlighted in green in Fig. 3.2.2), linear regression models are the subject of the analyses of the present paper: between visible driving scene contents and mental workload (Model A), between mental workload and eye measures (Model B) and between visible driving scene contents and eye measures (Model C). Present aims thus include whether and which factors, as previously introduced in the above literature, might be reliably related between:

- (1) Driving scene components (road curvature, traffic, signage, buildings, road surface, and vegetation)
- (2) Driver mental workload (subjective perceived effort)
- (3) Driver eye measures (pupil size, fixation duration, saccade amplitude)

The overall purpose of the present study can be conceived as a singular research question:

- (1) During returning attention to a driving task prior to taking control, what do the eyes look like from a person who thinks/feels that the driving will be more/less difficult or easy?

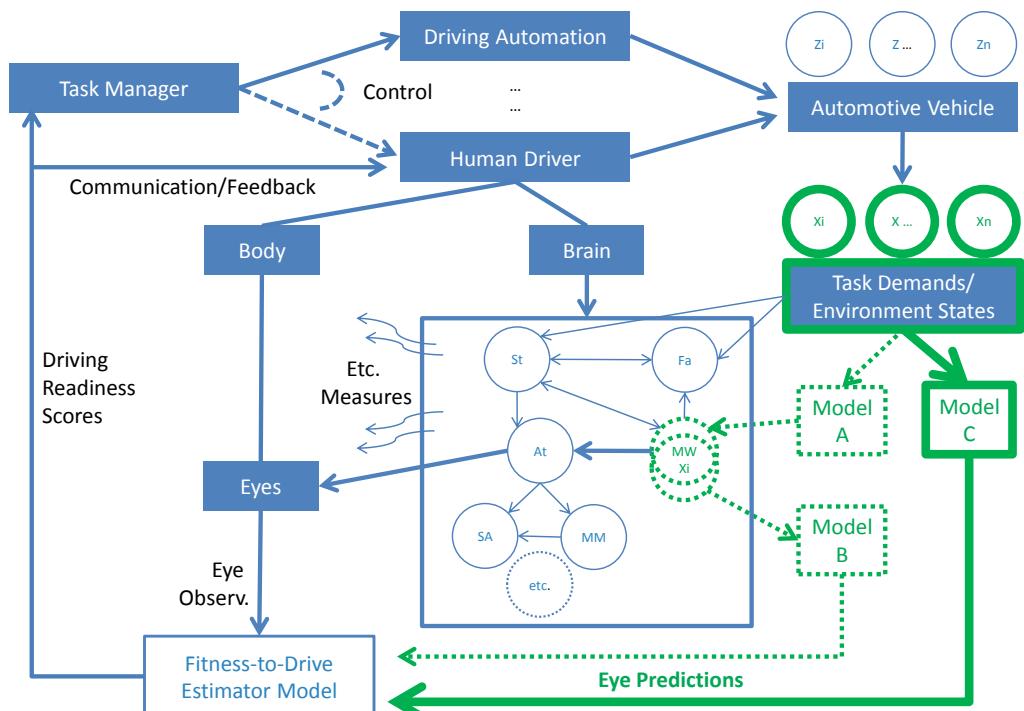


Figure 3.2.2. Based on a current driver's eye measures and reference eye predictions, a fitness-to-drive estimator model can supply driving readiness scores as communication/feedback or input to a task manager (entity or process) that can modulate transitions of driving control between human and automated agents. The overall design is modeled after a

framework of adaptive automation in Wickens & Hollands (2000, fig. 13.14, p. 547). The interior brain box is modeled after directional relations of several cognitive constructs from Heikoop et al. (2015, fig. 3, p. 9) where MW = Mental Workload, Fa = Fatigue, St = Stress, At = Attention, MM = Mental Models, and SA = Situation Awareness, while allowing for etc. = other mental states. Various computational models (A, B, C) are proposed and investigated within the present paper to predict eye measures from driving environment states (X_1, X_2, \dots, X_n) directly and/or in combination with Mental Workload.

2. Methods

We implemented an empirical approach to investigate the effects of various visible driving scene characteristics on human perceived effort ratings and corresponding eye behavior.

2.1 Participants and apparatus

Written informed consent was obtained from all participants, and the research was approved by the Human Research Ethics Committee of the Delft University of Technology under the title 'Driving video ratings' (16 December 2015). The experiment was completed by 15 participants (six female, nine male) aged between 18 and 36 ($M = 26.60$, $SD = 4.26$) with an average driving experience of around seven years since obtaining the driver's license ($M = 7.20$, $SD = 4.20$).

The experiment apparatus consisted of an isolating partition, a stimulus display monitor, eye tracker camera with integrated IR source and dedicated head/chin rest mount, as well as a gaming steering wheel (Fig. 3.2.3). The display was a 24 inch (diagonal) BenQ XL2420T-B monitor with a resolution of 1920 x 1080 pixels and a display area of 531 x 298 mm. The display was positioned about 95 cm in front of the participant and about 35 cm behind the eye tracking camera/IR source. The boundaries of the stimulus display area subtended approximately 31/18 degrees of horizontal/vertical viewing angle per the setup ranges required by guidelines of the SR Research Eyelink 1000 Plus eye tracker. Eye behavior data were recorded after individual participant calibration. The eye event parser was set according to the default psychophysical configuration recommended by the manual for research containing smooth pursuit movements and containing measurements of saccadic amplitude: saccade velocity threshold of 22 deg/s, saccade acceleration threshold of 3800 deg/s², and saccade motion onset delay threshold of 0 deg.

The gaming steering wheel was a Logitech G27 but was not connected to anything and along with the isolating partition was used to facilitate driving video stimulus immersion. Participants made use of a standard USB desktop mouse to input effort ratings on the stimulus display monitor.

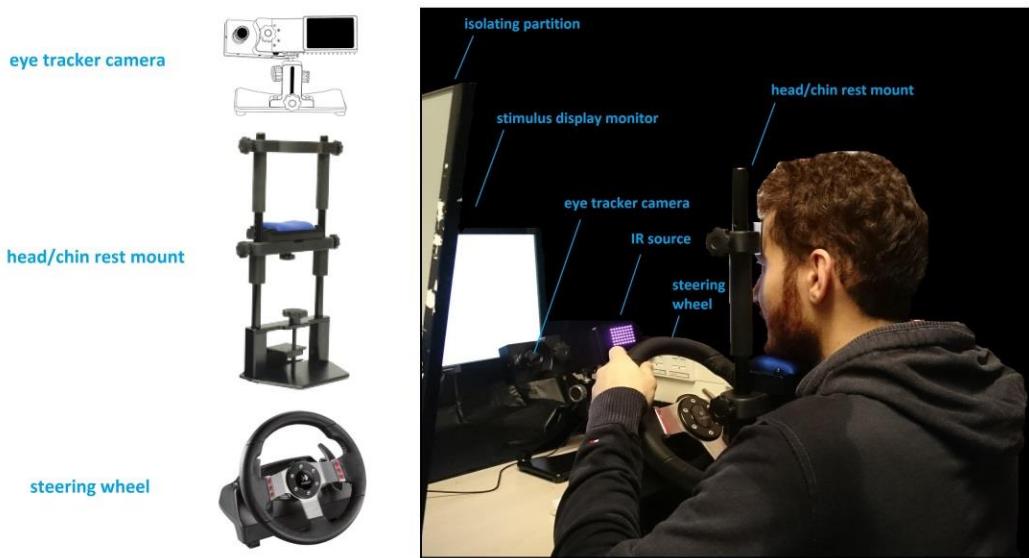


Figure 3.2.3. Diagram of experiment apparatus components and arrangement.

2.2 Procedure

Participants were encouraged to sit up straight and the height of the head/chin rest mount was adjusted to each participant to reduce potential neck/shoulder strain. Participants kept their heads stationary within the mount throughout the experiment except for voluntary rest breaks made available to them around every five minutes across about 15 minutes of driving video viewing and rating trials. Each trial began with an online drift correction dot in the center of the screen to which participants needed to fixate and click the mouse at the same time to begin. A 3 second long driving video clip was then played during which participants were tasked to move their hands to the wheel while imagining that they were taking over control (i.e., from automated driving) and that they must drive within that scene.

2.3 Stimuli and measurements

Stimuli consisted of a randomly ordered set of 60 dash cam video clips (Fig. 3.2.4) each of 3 seconds duration selected across a multi-level grouping extended from a semantic content categorization scheme developed in Cabrall et al., (2018) to ensure a variety of different driving scene circumstances. The duration of 3 seconds is on par with an average human reaction time for taking back control from an automated driving system found across a review of 25 papers in Eriksson and Stanton (2017). Driving scene content components were manually outlined and color coded (cf. CityScapes Dataset 2018) into five separate categories: (1) road surface area, (2) actual/potential road users (vehicles, bicycles, pedestrians), (3) signs/symbols (stop signs, cross walks, roadway writing, billboard advertisements, etc.), (4) buildings/infrastructure (houses, light poles, fences, etc.), and (5) vegetation (trees, bushes, hedges, etc.). A free online image processing tool (Krywinski, 2018) was used to determine a percentage of the windshield view that a specific category covered in terms of pixelated area. A transparent protractor overlay was used to approximate the curvature of the road of travel from the present lane position of the vehicle to the furthest distance point down the road.

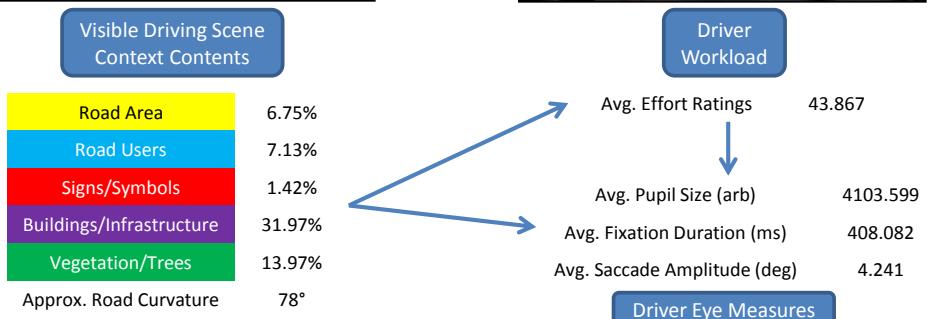
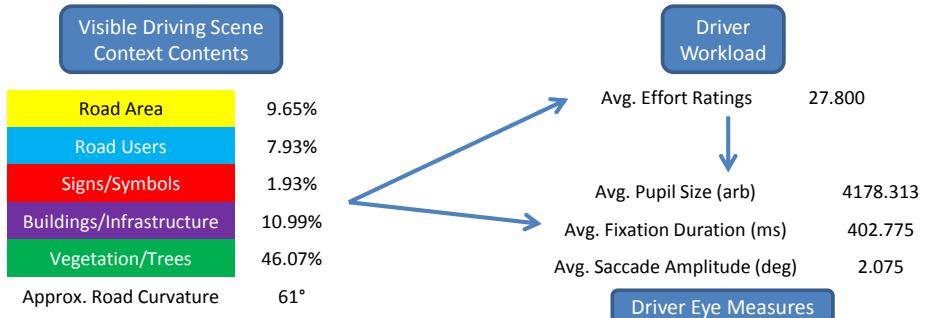
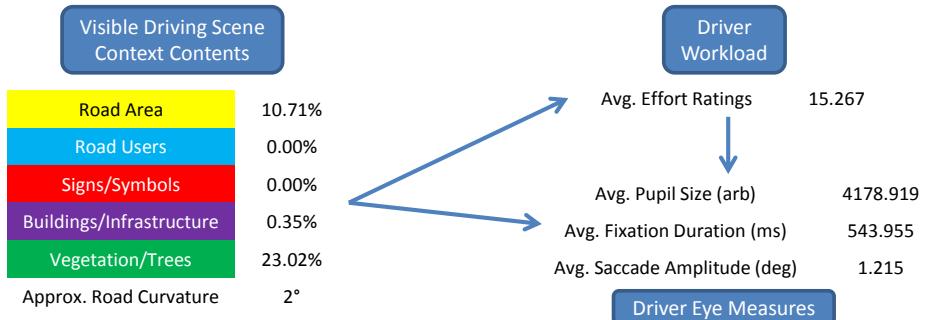


Figure 3.2.4. Segmentation of a driving video into quantifiable factor component predictors (road area, road users, signs/symbols, buildings/infrastructure, vegetation/trees, road curvature) and regression paths to outcome variables of workload (effort ratings) and eye measures (pupil size, fixation duration, saccadic amplitude).

All data were recorded to be analyzed at the level of a single driving video clip. For driving scene content description data (degree of road curvature and amounts of categorized pixel area coverage), a single representative frame was selected from the approximate middle of the video clip. All eye measures were likewise averaged across the entire video clip duration. Lastly, effort ratings were recorded from each participant for each video clip before being averaged across all participants ($n = 15$) for each one of the 60 different video clips.

After the clip finished and disappeared, an effort rating scale ("How much effort for you to take control and drive within that segment?") was presented on the upper half of the screen, and participants moved a vertical mouse cursor to click on the scale to input their answer from between "Very Low" to "Very High". Cursor click horizontal positions were divided by the pixel length of the scale and rounded to a single point resolution from 0 to 100. The presented horizontal effort scale contained 21 equally spaced demarcations from left to right following from those described within the seminal NASA TLX (Task Load Index) subscales (Hart & Staveland, 1988) and widely adopted across driver workload assessment research (see Fig. 3.2.5).

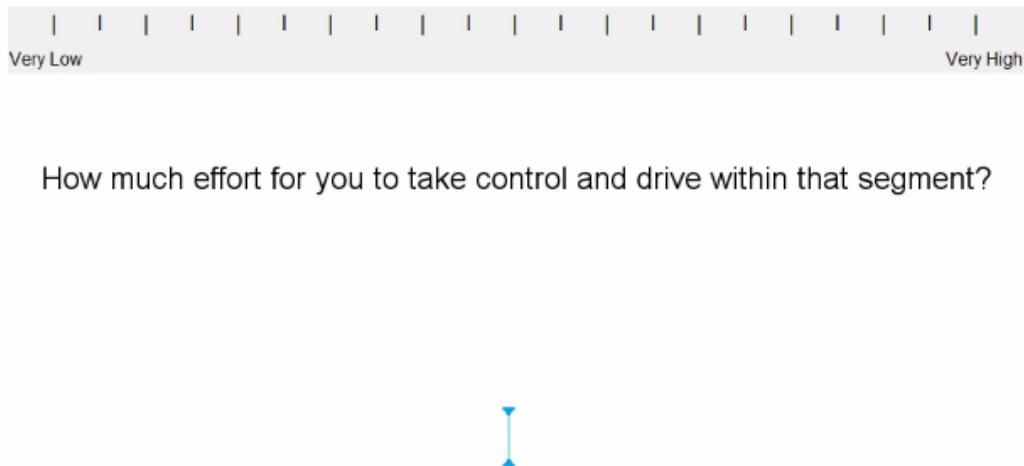


Figure 3.2.5. Driving effort response scale and cursor used to position on top of scale.

3. Results

Multiple rounds of linear regression models were applied to ascertain predictive power relations between several sets of independent variables (IV) and dependent variables (DV):

- 3.1. Model A: Driving scene contents (IV) and driver workload effort ratings (DV)
- 3.2. Model B: Driver workload effort ratings (IV) and driver eye measures (DV)
- 3.3. Model C: Driving scene contents (IV) and driver eye measures (DV)

First, pairwise correlations between the driving scene content independent variables were conducted to test for the presence of multicollinearity. All correlations between predictors were found to be well below a conventionally considered problematic threshold of $r = 0.80$ (Table 3.2.2, top). Second, the correlations of each driving scene content component with driver workload effort ratings and the eye measures were computed (Table 3.2.2, bottom). Furthermore, correlations and linear best fit lines for eye measures and scene content/characteristic by an index of video driving difficulty (wherein videos were grouped in accordance with the top, middle, and bottom third of ranked average video effort ratings) are depicted in Figure 3.2.6. Approximated degree of road curvature and pixelized area coverage amounts of signage, road users, and buildings/infrastructure all evidenced significantly positive correlations with driver workload effort ratings. None of the driving scene content components were found to significantly correlate with the pupil-size measure. Road curvature and signage showed significantly negative correlations with fixation durations. Road curvature, road users, buildings, and signage evidenced significantly positive correlations with saccade amplitude. For each eye measure, road curvature showed the highest correlative association.

Table 3.2.2. Correlations between driving scene contents, workload effort ratings, and driver eye measures. $N = 60$ video segments.

IVs ↓, IVs →	Road-Curve	Road-Users	Buildings	Road-Surface	Signs	Trees-Vegetation
Road-Curve	1	-	-	-	-	-
Road-Users	0.35*; ($p = 0.006$)	1	-	-	-	-
Buildings	0.21; ($p = 0.117$)	0.45*; ($p < 0.001$)	1	-	-	-
Road-Surface	0.02; ($p = 0.906$)	-0.18; ($p = 0.157$)	0.31*; ($p = 0.016$)	1	-	-
Signs	0.41*; ($p = 0.001$)	0.51*; ($p < 0.001$)	0.41*; ($p = 0.001$)	0.20; ($p = 0.129$)	1	-
Trees-Vegetation	-0.01; ($p = 0.959$)	-0.08; ($p = 0.545$)	-0.09; ($p = 0.479$)	-0.06; ($p = 0.660$)	-0.07; ($p = 0.616$)	1
DVs ↓, IVs →	Road-Curve	Road-Users	Buildings	Road-Surface	Signs	Trees-Vegetation
Workload-Effort	0.66*; ($p < 0.001$)	0.53*; ($p < 0.001$)	0.50*; ($p < 0.001$)	0.18; ($p = 0.169$)	0.55*; ($p < 0.001$)	0.06; ($p = 0.616$)
Pupil-Size	0.19; ($p = 0.158$)	-0.05; ($p = 0.698$)	-0.19; ($p = 0.153$)	-0.18; ($p = 0.172$)	-0.12; ($p = 0.376$)	0.24; ($p = 0.070$)
Fixation-Duration	-0.55*; ($p < 0.001$)	-0.25; ($p = 0.051$)	-0.20; ($p = 0.118$)	0.22; ($p = 0.093$)	-0.43*; ($p = 0.001$)	0.19; ($p = 0.144$)
Saccade-Amplitude	0.67*; ($p < 0.001$)	0.61*; ($p < 0.001$)	0.41*; ($p = 0.001$)	0.05; ($p = 0.696$)	0.59*; ($p < 0.001$)	-0.20; ($p = 0.126$)

* Correlation is significant at the 0.05 level (2-tailed)

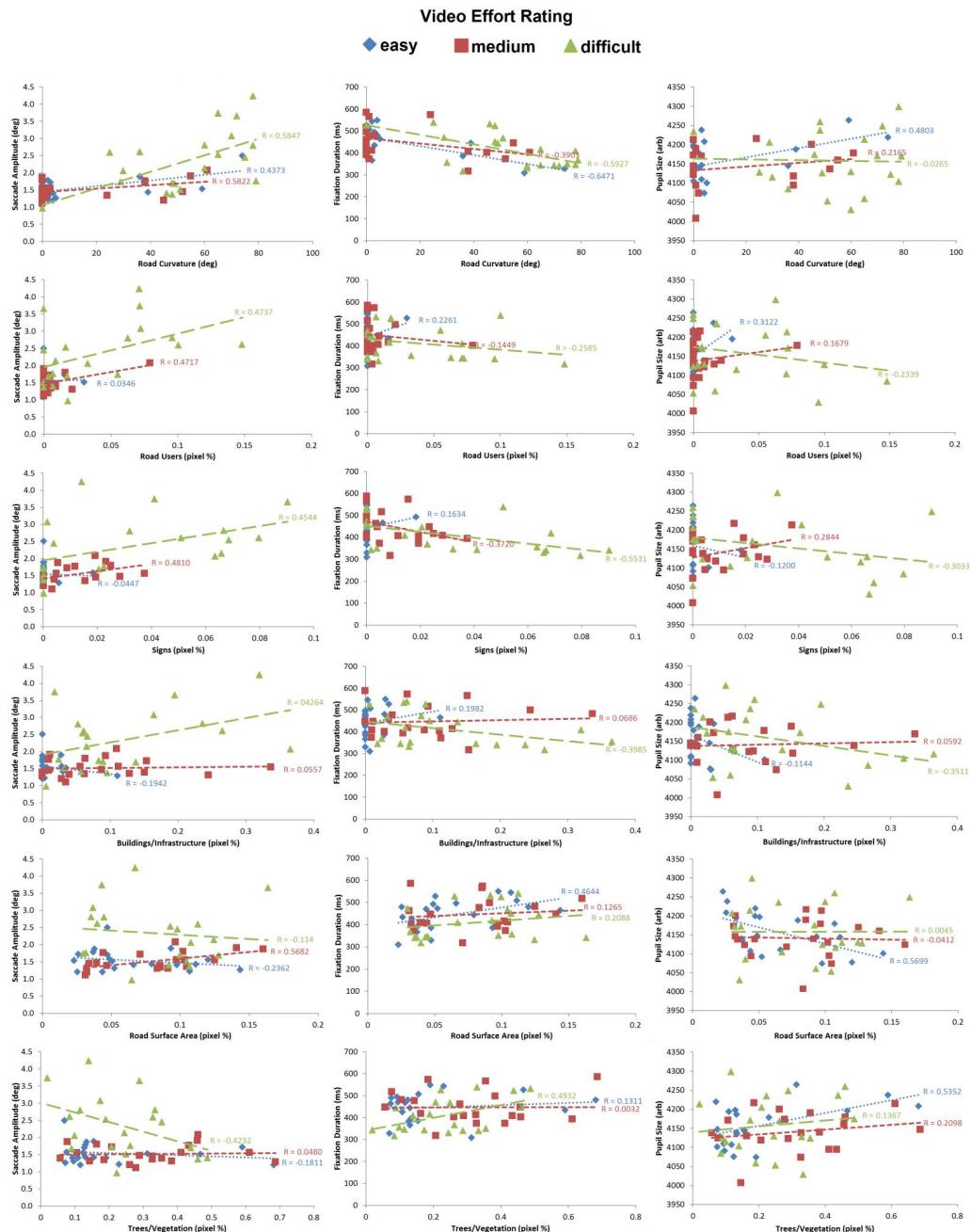


Figure 3.2.6. Correlations between independent variable driving scene content or characteristic ('x') and dependent eye measure ('y') by ranked video effort rating grouping (easy, medium, difficult).

3.1 Prediction of driver workload from driving scene contents

A multiple linear regression was conducted with all of the driving scene content component factors (road curvature, road users, buildings/infrastructure, road surface, signs/symbols, and vegetation/trees) entered as the predictor variables and with driver workload (effort ratings) as the outcome variable. The resulting equation (Eq. 1) was found to be statistically significant ($F(6,53) = 15.85$, $p < 0.001$) indicating that the combined driving scene factors taken together significantly improved the prediction of workload compared to the intercept model alone. Model summary statistics indicated around 64% of the variance in workload rating response scores were accounted for by the full set of driving scene content factors with a standardized error estimate of 7.6 (Table 3.2.3). The predictor factors of road curve angle, road users, and buildings/infrastructure were found to provide significant individual contribution to the amount of explained variance while controlling for the other predictor variables. Restriction of the model to only these predictors produced a lower variance accounted for (around 60%) and higher standard error estimate (around 7.8).

Table 3.2.3. Model summary statistics for prediction of driver workload from driving scene contents.

Model (Eq. 1)	$\hat{Y} = 10.398 + 0.202X_i + 90.232X_{ii} + 27.897X_{iii} + 42.528X_{iv} + 64.266X_{v} + 8.646X_{vi}$ $F(6,53) = 15.85$, $p < 0.001$			
	<i>r</i>	<i>r</i>²	<i>r</i>² adjusted	σ_{est}
	0.80	0.64	0.60	7.6
Predictor(s)	<i>Std. Error</i>	<i>B</i>	<i>t</i>	<i>p</i>
X_i Road-Curve	0.039	0.480	5.232	0.000*
X_{ii} Road-Users	43.648	0.235	2.067	0.044*
X_{iii} Buildings	13.749	0.210	2.029	0.047*
X_{iv} Road-Surface	30.964	0.134	1.373	0.175
X_v Signs	64.266	0.123	1.158	0.252
X_{vi} Trees-Vegetation	8.646	0.118	1.426	0.160

3.2 Prediction of driver eye measures from driver workload

Several linear regressions were conducted with driver workload (effort ratings) entered as the predictor variable with each eye measure (pupil size, fixation duration, saccade amplitude) taken in turn as a single outcome variable. Of the resulting equations (Eq. 2, Eq. 3, and Eq. 4) only those for fixation duration ($F(1,58) = 4.94$, $p = 0.030$) and saccade amplitude ($F(1,58) = 34.66$, $p < 0.001$) were found to be statistically significant, while significant difference was not obtained for pupil size ($F(1,58) = 1.08$, $p = 0.303$). Less than 1% of the variance in pupil size was found to be explainable from the workload ratings (Table 3.2.4). Model summary statistics indicated around 6% of the variance in fixation durations was accounted for by the workload ratings with a standardized error estimate of 71.3 ms (Table 3.2.5), and that around 36% of the variance in saccade amplitude was accounted for by the workload ratings with a standardized estimate of 0.5 degrees (Table 3.2.6).

Table 3.2.4. Model summary statistics for prediction of driver pupil size from driver workload.

Model (Eq. 2)	$\hat{Y} = 4134.718 + 0.682Xi$ $F(1,58) = 1.08, p = 0.303$
	r r^2 $r^2 \text{ adjusted}$ σ_{est}
	0.14 0.02 0.001 60.6
Predictor(s)	Std. Error <i>t</i> <i>p</i>
X_i Workload-Effort	0.656 0.135 1.040 0.303

Table 3.2.5. Model summary statistics for prediction of driver fixation durations from driver workload.

Model (Eq. 3)	$\hat{Y} = 478.973 + -1.717Xi$ $F(1,58) = 4.94, p = 0.030$
	r r^2 $r^2 \text{ adjusted}$ σ_{est}
	-0.28 0.08 0.06 71.3
Predictor(s)	Std. Error <i>t</i> <i>p</i>
X_i Workload-Effort	0.773 -0.280 -2.222 0.030*

Table 3.2.6. Model summary statistics for prediction of driver saccade amplitude from driver workload.

Model (Eq. 4)	$\hat{Y} = 0.929 + 0.34Xi$ $F(1,58) = 34.66, p < 0.001$
	r r^2 $r^2 \text{ adjusted}$ σ_{est}
	0.61 0.38 0.36 0.5
Predictor(s)	Std. Error <i>t</i> <i>p</i>
X_i Workload-Effort	0.006 0.612 5.888 < 0.001*

3.3 Prediction of driver eye measures from driving scene contents

Several multiple linear regressions were conducted with all of the driving scene content component factors (road curvature, road users, buildings/infrastructure, road surface, signs/symbols, and vegetation/trees) entered as the predictor variables with each eye measure (pupil size, fixation duration, saccade amplitude) taken in turn as a single outcome variable. Of the resulting equations (Eq. 5, Eq. 6, and Eq. 7) only those for fixation duration ($F(6,53) = 9.216, p < 0.001$) and saccade amplitude ($F(6,53) = 18.171, p < 0.001$) were found to be statistically significant, while significant difference was not obtained for pupil size ($F(6,53) = 1.681, p = 0.144$). Around 7% of the variance in driver pupil size was found to be explainable from the driving scene content factors (Table 3.2.7). Model summary statistics indicated around 46% of the variance in driver fixation durations was accounted for by the driving scene content factors with a standardized error estimate of 54.4 (Table 3.2.8), and around 64% of the variance in driver saccade amplitude was accounted for by the driving scene content factors with a standardized error estimate of 0.4 (Table 3.2.9). For fixation durations, the predictor factors of road curve angle, road surface, signs, and road users were found to provide significant individual contribution to the amount of explained variance while controlling

for the other predictor variables. For saccade amplitudes, the predictor factors of road curve angle and road users were found to provide significant individual contributions to the amount of explained variance while controlling for the other predictor variables.

Table 3.2.7. Model summary statistics for prediction of driver pupil size from driving scene contents.

Model (Eq. 5)	$\hat{Y} = 4142.455 + 0.588X_i + 77.18X_{ii} + -182.965X_{iii} + -335.847X_{iv} + -81.233X_v + -63.807X_{vi}$ $F(6,53) = 1.68, p = 0.144$			
	<i>r</i>	<i>r</i> ²	<i>r</i> ^{2 adjusted}	σ_{est}
	0.40	0.16	0.07	58.6
Predictor(s)	Std. Error	<i>B</i>	<i>t</i>	<i>p</i>
X_i Road-Curve	0.299	0.277	1.970	0.054
X_{ii} Trees-Vegetation	46.859	0.209	1.647	0.105
X_{iii} Road-Surface	239.220	-0.144	-0.765	0.448
X_{iv} Signs	428.773	-0.128	-0.783	0.437
X_v Buildings	106.217	-0.121	-0.765	0.448
X_{vi} Road-Users	337.210	-0.033	-0.189	0.851

Table 3.2.8. Model summary statistics for prediction of driver fixation duration from driving scene contents.

Model (Eq. 6)	$\hat{Y} = 400.83 + -1.179X_i + 824.614X_{ii} + -1196.586X_{iii} + 653.48X_{iv} + -158.252X_{v} + 86.068X_{vi}$ $F(6,53) = 9.22, p < 0.001$			
	<i>r</i>	<i>r</i> ²	<i>r</i> ^{2 adjusted}	σ_{est}
	0.72	0.51	0.46	54.4
Predictors	Std. Error	<i>B</i>	<i>t</i>	<i>p</i>
X_i Road-Curve	0.277	-0.457	-4.257	< 0.001*
X_{ii} Road-Surface	221.958	0.423	3.715	< 0.001*
X_{iii} Signs	397.832	-0.375	-3.008	0.004*
X_{iv} Road-Users	312.876	0.278	2.089	0.042*
X_v Buildings	98.553	-0.194	-1.606	0.114
X_{vi} Trees-Vegetation	43.478	0.192	1.980	0.053

Table 3.2.9. Model summary statistics for prediction of driver saccade amplitude from driving scene contents.

Model (Eq. 7)	$\hat{Y} = 1.395 + 0.011X_i + 6.471X_{ii} + 5.673X_{iii} + -0.612X_{iv} + 0.571X_v + 0.512X_{vi}$ $F(6,53) = 18.17, p < 0.001$			
	<i>r</i>	<i>r</i> ²	<i>r</i> ^{2 adjusted}	σ_{est}
	0.82	0.67	0.64	0.4
Predictors	Std. Error	<i>B</i>	<i>t</i>	<i>p</i>
X_i Road-Curve	0.002	0.463	5.280	0.000*
X_{ii} Road-Users	2.318	0.304	2.791	0.007*
X_{iii} Signs	2.948	0.196	1.925	0.060
X_{iv} Trees-Vegetation	0.322	-0.150	-1.900	0.063
X_v Buildings	0.730	0.077	0.782	0.438
X_{vi} Road-Surface	1.645	0.029	0.312	0.757

4. Discussion

4.1 Using driving scenes to predict driver workload

The design of our experiment exposed participants to multiple driving scenes, which varied in the amount of 'stuff' visible in the scene, and a range of experienced workload effort ratings was captured. Specifically, the average workload ratings per video ranged between 4.67 and 55.87 on the provided 0 to 100 point scale. As expected, this indicates that our non-critical/non-hazardous driving scene stimuli represented a range of perceived difficulties even while collectively residing on the lower portion of a scale for anticipated driving effort. Moreover, the workload ratings were found to vary in a reliable way in accordance with the identified driving scene components, accounting for approximately 60.2% of the variance of the workload effort ratings which can be interpreted as a '*moderately strong*' relation in accordance with the 5-level general guide provided by Brewer (2003). A standard error estimate of around 7 points on a 100 point workload scale is expected to be practically useful (even without being perfectly precise) in adaptive aiding and driver monitoring systems, for example as depicted within Fig. 3.2.2.

Such results are consistent with general practices of road safety research to factorize driving task scenarios into more/less easy conditions, and in particular are convergent with the results of Steyver et al. (1994) where differently experienced appreciations were found by tasking participants to imagine having to drive on various roads via use of previously recorded video footage. For causal effects from scene components to workload, the previously introduced and reviewed research has been purposefully constrained to a limited number of conditions with controlled variations in combinations of different hypothesized factors of interest (e.g., day/night, presence/absence of traffic, homogeneous/heterogeneous vegetation, presence/absence highway merging, presence/absence of roadside barriers, etc.) and thus participants had been typically exposed to only a handful of different road scenes. As an extension to such research, and through use of modern day image processing technology, the present computational approach was able to employ 60 different videos with continuous rather than discrete quantities of scene components while collecting human judgments of effort on a continuous scale inspired by the standardized and widely adopted NASA-TLX workload measure.

Not all visual information in the scene was found to be correlated with the effort rating responses (i.e., trees/vegetation and road surface were not statistically significant) ahead of the regression model. The driving scene features and objects that were found to be significantly correlated with effort rating responses (i.e., amounts of road curvature, road users, signs, and buildings) appear as a group to be those that are perhaps more semantically meaningful to the task of receiving driving control. While increased amounts of buildings and trees both present additional obstacles a driver must be wary of avoiding, the former is conceivably more likely to entail presence of other people and vehicles and hence signs/symbols for negotiating rules governing their interactions. Within the regression model, it remains to be seen why specifically the individual contribution of signage/symbols to driving effort ratings is reduced given the presence of the predictors. For now, a plausible interpretation is that signs not only co-exist with, but are often designed specifically as warnings for curved roads, potential presence of other road users, and identification of buildings, etc. Such a notion is supported by our observed co-variance matrix (Table 2, top) and hence why the individual explanatory power of signage might be diminished given the presence of the other significantly contributing features.

4.2 Using driver workload to predict driver eye measures

Given first considerations of the previously reviewed research on driver eye measure correlates of workload, it would appear as though our eye measurement results are lacking at best and contradictory at worse. From increases in workload effort ratings: pupil sizes did not systematically increase, fixations durations were longer rather than shorter, and saccade amplitudes increased rather than decreased. Limitations for not obtaining increased pupil sizes with increased workload in our study might be considered as stemming from a lack of strict measurement and control over luminance effects from the varying driving video clips (e.g., some might have been systematically brighter/dimmer than others irrespective of effort related factors). The pupil is well-known to be more dynamically susceptible in terms of size changes due to adaptive responses to lighting rather than cognitive states. Indeed, pupillometry in driving applications has been criticized by such real world limitations as recognized for vehicles that physically traverse natural environments of varying light and shadow, etc. without recompense to increased differentiating resolutions from sophisticated compensatory technology and/or costly patented algorithms.

However, more reasonable and practical interpretations of our seemingly problematic results are derivable following a similar argument logic as used by Recarte et al (2008) and Gerhard et al. (2015) to rectify the mixed results regarding driver workload and driver blink rates (cf. Kramer, 1990). The mixed results regarding blink rate (cf. Kramer, 1990) have previously been deemed explainable due to situational aspects (i.e., visual demands) and the theoretical (non)differentiation of such confounds whereby visual and mental workload may produce eye measure results in opposite directions (Recarte et al., 2008; Marquart et al., 2015). Furthermore, consensus interpretations could be better served by direct consideration of what is expected to be an intrinsic or extrinsic component of nominal primary driving task demands. For example, Recarte et al., 2008 found an increase in blink rate during all their secondary cognitive tasks (listening, talking, and calculating). Hence, it might be initially generalized and expected for example that more complex driving scenarios like urban areas should be associated with increased blink rates. However, Recarte et al. 2008 also found a decrease in eye blink rate for more visually demanding tasks when compared to less visually demanding tasks. Now another perspective might expect, for example, dense urban areas to entail increased visual demands such as from traffic lights, road signs, road markings, and buildings, etc. and thus serve to reduce blink rate as the observer prolongs looking exposures to take in and process the large amount of visual information.

Reasonably, mental driving effort does not necessarily follow from an increase in only visual information, *per se*, as drivers may not, in fact, be tasked to take in *all* information but instead prioritize along subsets of the most meaningful aspects (e.g., as relevant to their driving task). In other words, not all information (whether visual or mental) necessarily presents itself as a demand (i.e., of primary driving task relevancy). Although other eye measures of mental workload have conventionally been interpreted to show more directional consistency than blink rate (Table 3.2.1), the application and interpretation of eye behavior data in driving may warrant reconsideration when taking into account such situated aspects (i.e., consideration and comparison of general visual information versus those specific to effortful mental driving demands). Of those previously introduced and reviewed research that found eye results in counter direction to our own at present, eye measurements were often taken under circumstances to purposefully induce extra mental workload or account for spare capacity via secondary tasks (e.g., auditory recall, cell phone conversations, etc.) and/or to elevate urgency as with safety-critical driving scenarios (e.g.,

hazards). Thus, a sourcing explanation and research focus for driver mental workload is deemed most appropriate at present.

In the recently updated psychological construct model for driving automation by Heikoop et al. (2015), mental workload can be attributed (among other interactive cognitive states) to be impacted by task demands before in turn impacting attention and eventually situation awareness. Logically, task demands for a driver can come from primary driving tasks as well as secondary tasks (e.g., driving unrelated). Additionally, primary task demands can both be considered across a full spectrum of nominal and off-nominal conditions. Our present research purposes were to computationally model and establish reference eye measures (presumably of mental workload) to inform assessments of a driver-getting-ready-to-drive under different (mathematically describable) driving scene circumstances. As a starting point, and for a probabilistically high proportion of impact to driving operations, nominal and non-critical transitions of driving control away from automation towards increased human driving responsibility were selected for investigation. In our modeling of non-critical levels of driving control workload effort within nominal driver eye measures (i.e., without additional secondary tasks or while evaluating hazards), we have found a lack of significance for our workload ratings to be positively associated with pupil sizes, while negative and positive relations reached significance for fixation durations and saccade amplitudes, respectively. Per the guidelines of Brewer (2003), the obtained correlative explained variances of around < 1%, 6%, and 36% for pupil size, fixation duration, and saccade amplitude might be appropriately interpreted as '*none*', '*weak*', and '*mild/modest*' respectively. The obtained standard error estimates of around 71 ms (for fixation durations) and half a degree of gaze angle (for saccade amplitude) are expected to be of sufficient resolution to be practically useful as predicted targets to represent some component of mental workload within the eye measures of would-be drivers.

4.3 Using driving scenes to predict driver eye measures

Irrespective of any precisely captured cognitive state, the design of our study was able to successfully capture systematic predictions of eyes movements as a function of different compositions of driving scenes. Generally, the correlative directions of amount of scene content impact on fixation duration (negative) and saccade amplitude (positive) are consistent with previous driver eye tracking research that separately controlled for road type visual complexity vs. aspects of danger. Chapman and Underwood (1998) found that when nominally safe road types are compared along a dimension of rising visual complexity such as with rural, suburban, and urban roads, then average fixation durations present a decreasing pattern of 437, 420, and 389 ms respectively, while average saccade lengths exhibit an increasing pattern of 1.71, 1.99, and 2.16 degrees respectively. While an opposing direction is presented by elevating the danger of a driving situation (increase in fixation durations, and decrease in saccade lengths), by comparing across road types within a collapsed condition of dangerous situations, then the same previous directional eye movement patterns persist again (decrease in fixation durations, and increase in saccade lengths) as a function of increasing roadway visual complexity.

However, as with the interpretation of our workload ratings in 4.1, not all visual information in the driving scene was found to be significantly correlated with the eye measure outputs of fixation durations and saccade amplitudes. Ahead of the regression models, only degree of road curvature and amount of signage/symbols were significantly correlated with fixation durations (both in a negative direction), whereas road curvature, road users, buildings, and amount of signs were

significantly correlated with saccade amplitudes (all in a positive direction). Conversely, an association in either direction of amount of road surface area and amount of vegetation did not obtain significant correlation with either fixation duration or saccade amplitude. In pretending to initially receive driving control and assess a scene '*in medias res*' for its required driving effort, the eyes of our participants appeared to be motivated to move more quickly and further around especially as related to increased degree of road curvature and presence of signage/symbols (as both items were significantly correlated for both fixation durations and saccade amplitude). Within the regression models, the strongest contributing factor predictive of either fixation duration or saccade amplitude eye movements was degree of road curvature. Approximately 7%, 46% and 64% of the variance explained in pupil size, fixation duration, and saccade amplitude from the driving scene component factors represents associations that by the labels of Brewer (2003) would be considered '*weak*', '*moderate*', and '*moderately strong*', respectively. Obtained standard estimates of around 54 ms (for prediction error of fixation durations) and around 0.4 degrees (for prediction error of saccade amplitude) are expected to be of sufficient resolution to be practically useful as eye movement targets to represent nominal driving scene evaluation processes.

As with the general scene viewing literature reviewed in our introduction, we expect a blend of exogenous and endogenous influence on eye movements. However, for our participants and the target application area of transitioning control to human drivers, the relevant visual sampling task is not an open one of free exploratory viewing or a closed one of searching for a particular item/feature but instead a purposeful time limited assessment is assumed to transpire during the transitional phase of getting ready to drive. Thus, the initial movements of the eyes before re-uptake of driving control are most likely driven both by scene stimulus saliency factors (amounts of visual information) but can quickly mediate and ignore irrelevant visual complexities (e.g., trees/vegetation) and mediate gaze patterns instead by semantic relevancy to specifics of driving task demands on a higher cognitive level regarding where/how to look around (reduction of lateral and longitudinal conflict risks; adherence to governing rule/regulations; and negotiation of potential interactions with other road users).

4.4 Potential Applications

As previously introduced, the most relevant application areas for the findings of the present investigation, its produced regression model equations, and the proposed adaptive control framework depicted in Fig. 3.2.2, are envisioned to be within the transitional phases between automated and human driving control as depicted in the middle of Fig. 3.2.1. Ahead of being given full control of a vehicle, the eyes of the would-be driver can be compared against stored predicted values as per the present situation of the given driving scene components. As the person him/herself begins to assess the driving scene and ascertain how much the task demands will require of their mental effort, attentional resources, and ultimately situation awareness, an automatic driver monitoring system safety layer can provide oversight and correct as needed. If the driving scene is one where moderate/high amounts of driving workload effort are expected (e.g., containing a sizeable amount of other road users, a large road curvature degree, many signs and symbols to read and interpret, etc.) but the driver's eyes are moving slower and with shorter distances than that has been previously computationally predicted (e.g., they are still mentally fixated on that last email they were composing), then any number of different adjustments might be made in terms of automated warnings and/or vehicular control to modulate the potential risks of the transition. One example interface solution might be to begin to highlight relevant missed parts of the driving scene until the driver is able to unlock full manual control by gazing at these,

but of course there are many alternative design solutions. In any case, a real-time assessment of driver fitness to drive within the present scene situation would be desirable and the present study represents a starting point method and resulting model for generating such information. Note, on account of all the models (A, B, and C) outlined in Fig. 3.2.2, the solution framework need not be limited to a strictly behavioristic or cognitive perspective but is amenable to either or a blend of both.

4.5 Limitations

There are important considerations in common across our measures that should be taken into account. All obtained effect sizes were taken into account only after an averaging of multiple exposures rather than a single video viewing where the effect sizes would be expectedly reduced. Additionally (due to difficulties in valid/reliable manual human annotation), a potentially confounding effect of ego-vehicle speed in the driving video segments was not yet controlled/characterized, and we recommend such an aspect as an interesting mediating or independent factor to investigate in future studies. Eye measurements were taken while viewing a previously filmed driving video rather than in full fidelity environment where additional fields of view might be expected to be present and relevant (e.g., mirrors and peripheral). Eye tracking and scene identification will be limited by the cost and availability of technological software and hardware components (e.g., computer vision and machine learning) although these have been recently undergoing rapid advancements. Additional model components of other non-intrusive physiological measures and cognitive

construct interactions would be expected to complement the present envisioned adaptive control model provided in Fig. 3.2.2. As with all models, more data is expected to improve the presently provided regression equations – the current framework could be extended via additional videos of greater variety and increased sets of classifiable items and/or greater resolutions of bounding boxes.

5. Conclusions

In conclusion, the present study contributes new regression model equations that are statistically significant improvements by including their identified predictor variable factors over simple intercept only counter-parts for each of the following relations of interest: for predicting driver mental workload from visible driving scene contents, for predicting driver eye movements from driver mental workload, and for predicting driver eye movements directly from driving scene contents. Such models are applicable during transitions of control away from automated driving that would involve an initial human viewing of a driving scene for the purpose of evaluating the amount of effort that might be needed for uptake of conventional manual driving control. Automated support can be designed in a variety of circumstances in cases of mis-matches between eyes of drivers and driving scene contents. For example, eyes measured as exhibiting too low saccadic amplitudes for too long in a driving scene of high complexity contents might indicate an unawares driver (a.k.a., “looking but not seeing”) thus suggesting a prolonged involvement of an automated driving agent if possible, or ultimately safe-stop procedure if available. On the other hand, eyes measured as over-expressing a nominal level of driving scene complexity in a more simplistic scene might be useful for facilitative training aids for novice or otherwise overwhelmed drivers.

Acknowledgments

The research presented in this paper was conducted within the project HFAuto – Human Factors of Automated Driving (PITN-GA-2013-605817).

References

Ahern, S.K. (1978). Activation and intelligence: Pupillometric correlates of individual differences in cognitive abilities. Doctoral dissertation, University of California.

BASt (2012). Definitions of Automation and Legal Issues in Germany. TRB Road Vehicle Automation Workshop. Irvine, CA, USA. <http://onlinepubs.trb.org/onlinepubs/conferences/2012/Automation/presentations/Gasser.pdf>

Beatty, J. (1982) Task-evoked pupillary responses, processing load, and the structure of processing resources. *Physiological Bulletin*, 91, p. 276–292.

Brewer, D. (2003). Introduction to Statistics: Regression and Correlation. Online course materials. <http://faculty.washington.edu/ddbrewer/s231/s231regr.htm>

Buswell, G.T. (1935). How People Look at Pictures. Chicago: University of Chicago Press.

Cabral, C.D.D., Lu, Z., Kyriakidis, M., Manca, L., Dijksterhuis, C., Happee, R., & De Winter, J.C.F. (2018). Validity and reliability of naturalistic driving scene categorization judgments from crowdsourcing. *Accident Analysis & Prevention*, 114, p. 25–33

Chapman, P.R., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception*, 27, p. 951–964

Cityscapes Dataset (2018). Semantic Understanding of Urban Street Scenes. <https://www.cityscapes-dataset.com/>

Cummings, M. L., & Guerlain, S. (2007). Developing operator capacity estimates for supervisory control of autonomous vehicles. *Human Factors*, 49, p. 1–15.

De Waard, D. (1996). The measurement of drivers' mental workload. Ph.D. Thesis. Traffic Research Centre, University of Groningen. Haren, The Netherlands.

De Winter, J.C.F., Eisma, Y.B., Cabral, C.D.D., Hancock, P.A., Stanton, N.A. (in press). Situation awareness based on eye movements in relation to the task environment. *Cognition Technology and Work*.

Durso, F.T., Bleckley, M.K., & Dattel, A.R. (2006). Does situation awareness add to the validity of cognitive tests? *Human Factors*, 48, p. 721–733.

Durso, F.T., Hackworth, C.A., Truitt, T.R., Crutchfield, J., Nikolic, D., & Manning, C. A. (1998). Situation awareness as a predictor of performance for en route air traffic controllers. *Air Traffic Control Quarterly*, 6, p. 1–20.

Eisma, Y.B., Cabral, C.D.D., De Winter, J.C.F. (2018) Visual sampling processes revisited: Replicating and extending Senders (1983) using modern eye-tracking equipment. *IEEE Transactions on Human-Machine Systems*, PP(99), p.1–15, DOI: 10.1109/THMS.2018.2806200 <https://ieeexplore.ieee.org/abstract/document/8331130/>

Endsley, M.R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 32, p. 97–101.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, 37(1), p. 65–84.

Engström, J., Monk, C.A., Hanowski, R.J., Horrey, W.J., Lee, J.D., McGehee, D.V., Regan, M., Stevens, A., Traube, E., Tuukkanen, M., Victor, T., & Yang, C.Y. (2013). A Conceptual Framework and Taxonomy for Understanding and Categorizing Driver Inattention. Brussels, Belgium: European Commission. <https://ec.europa.eu/digital-single-market/news/us-eu-inattention-taxonomy-report>

Eriksson, A. & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors*, 59(4), p. 689–705.

Eyeris (2018). Interior Vision AI for Occupants of Highly Automated Vehicles. <http://www.eyeris.ai/>

eyeSight (2018). Enjoy a Better and Safer Driving Experience with eyeSight. <http://www.eyesight-tech.com/product/automotive/>

Gardner, A.K., Kosemund, M., & Martinez, J. (2017). Examining the feasibility and predictive validity of the SAGAT tool to assess situational awareness among surgical trainees. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 12, p. 17–21.

Gibson, J.J., and L.E. Crooks (1938). A theoretical field-analysis of automobile-driving. *The American Journal of Psychology*, 51(3), p. 453–471.

Gugerty, L. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied*, 3, p. 42–66.

Gutzwiller, R.S., & Clegg, B.A. (2013). The role of working memory in levels of situation awareness. *Journal of Cognitive Engineering and Decision Making*, 7, p. 141–154.

Hart, S., and Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52, p. 139–183.

Heikoop, D., De Winter, J.C.F., van Arem, B., Stanton, N.A. (2015). Psychological constructs in driving

automation: A consensus model and critical comment on construct proliferation. *Theoretical Issues in Ergonomics Science*, 17(3), p. 284–303.

Helmholtz, H. V. (1910). Hanbuck der Physiologischen Detik. English translation, J. Southall. Dover: New York, 1925.

Henderson, J.M. & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.) *Eye Guidance in Reading and Scene Perception*. Oxford: Elsevier. p. 269-293.

Ikuma, L. H., Harvey, C., Taylor, C. F., & Handal, C. (2014). A guide for assessing control room operator performance using speed and accuracy, perceived workload, situation awareness, and eye tracking. *Journal of Loss Prevention in the Process Industries*, 32, p. 454–465.

Just, M.A., & Carpenter, P.A. (1980). A theory of reading: from eye fixation to comprehension. *Psychological Review*, 87, p. 329–354.

Kircher, K., & Ahlstrom, C. (2016). Minimum required attention: a human centric approach to driver inattention. *Human Factors*, 59(3), p. 471–484.

Kircher K., & Ahlstrom, C. (2018) Evaluation of methods for the assessment of attention while driving. *Accident Analysis and Prevention*, 114, p. 40–47.

Kramer, A.F. (1990). Physiological metrics of mental workload: A review of recent progress. In D. Damos (Ed.), *Multiple-task Performance*. London: Taylor & Francis, p. 279–328.

Krzywinski, M. (2018). Image color summarizer: RGB, HSV, LCH & Lab Image Color Statistics and Clustering – Simple and Easy. <http://mkweb.bcgsc.ca/color-summarizer/?analyze>

Lee, J.D. (2008). Fifty years of driving safety research. *Human Factors*, 50, p. 521–528.

Lo, J.C., Sehic, E., Brookhuis, K.A., & Meijer, S.A. (2016). Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers. *Transportation Research Part F: Traffic Psychology and Behavior*, 43, p. 325–338.

Loft, S., Bowden, V., Braithwaite, J., Morrell, D. B., Huf, S., & Durso, F. T. (2015). Situation awareness measures for simulated submarine track management. *Human Factors*, 57, p. 298–310.

Loftus, E., Loftus, G., & Messo, J. (1987). Some facts about "Weapon Focus". *Law and Human Behavior*, 11(1), p. 55-62.

Lu, Z., Coster, X., & De Winter, J.C.F. (2016) How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics*, 60, p. 293–304.

Lu, Z., Happée, R., Cabrall, C.D.D., Kyriakidis, M., & De Winter, J.C.F. (2016). Human factors of transitions in automated driving: A general framework and literature survey. *Transportation Research Part F: Traffic Psychology and Behavior*, 43, p. 183–193.

Maltz, M., & Shinar, D. (1999). Eye movements of younger and older drivers. *Human Factors*, 41(1), p. 15–25.

Marquart, G., Cabrall, C.D.D., De Winter, J.C.F. (2015). Review of eye related measures of drivers' mental workload. *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*, *Procedia Manufacturing*, 3, p. 2854–2861.

Marquart, G. & De Winter, J.C.F. (2015). Workload assessment for mental arithmetic tasks using the task evoked pupillary response. *PeerJ Computer Science*, 1, e16.

Marshall, S. (2000). Method and Apparatus for Eye Tracking and Monitoring Pupil Dilation to Evaluate Cognitive Activity. U.S. patent no. 6,090,051.

Marshall, S.P. (2002). The index of cognitive activity: Measuring cognitive workload. *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*. Scottsdale, AZ, USA.

Mathworks. (2018). Computer Vision System Toolbox: Design and Simulate Computer Vision and Video Processing Systems. <https://www.mathworks.com/products/computer-vision.html>

May, J.G., Kennedy, R.S., Williams, M.C., Dunlap, W.P., & Brannan, J.R. (1990). Eye movement indices of mental workload. *Acta Psychologica*, 75(1), p. 75–89.

McGowan, A., & Banbury, S. (2004). Interruption and reorientation effects of a situation awareness probe on driving hazard anticipation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48, p. 290–294.

Michon, J.A. (1979). Dealing with Danger. Technical Report no. VK 79-01. Traffic Research Centre of the University of Groningen, The Netherlands.
http://www.jamichon.nl/jam_writings/1979_dealing_with_danger.pdf

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*. New York:Plenum Press, p. 458–520.
http://jamichon.nl/jam_writings/1985_criticial_view.pdf

Miura, T. (1990). Active functions of eye movement and useful field of view in a realistic setting. In R. Groner, et al. (Eds.), *From Eye to Mind*. Oxford, England: North-Holland Publishing, p. 119–127.

Moray, N. (1990). Designing for transportation safety in the light of perception, attention, and mental models. *Ergonomics*, 33, 1201–1213.

Moray, N. (1993). Designing for attention. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, Awareness, and Control*. Oxford, England: Clarendon Press, p. 111–134.

Mok, B., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., & Ju, W. (2015). Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles. *Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems*.

NHTSA (2017). Automated Driving Systems 2.0: A Vision for Safety. Washington, D.C.: National Highway Transportation Safety Administration.

NHTSA (2013). Preliminary Statement of Policy Concerning Automated Vehicles. Washington, DC.: National Highway Traffic Safety Administration.

O'Brien, K.S., & O'Hare, D. (2007). Situational awareness ability and cognitive skills training in a complex real world task. *Ergonomics*, 50, p. 1064–1091.

Ohn-Bar, E., & Trivedi, M.M. (2016). Looking at humans in the age of self-driving and highly automated vehicles. *IEEE Transactions on Intelligent Vehicles*, 1, p. 90–104.

Palinko, O., Kun, A., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. *Proceedings of the Eye Tracking Research and Applications Conference*

Papelis, Y., Ahmad, O. & Watson, G. (2003). Developing scenarios to determine effects of driver performance techniques for authoring and lessons learned. *Proceedings of the Driving Simulation Conference North America*. Dearboarn, Michigan, USA.

Petermeijer, S.M., De Winter, J.C.F., & Bengler, K. (2016). Vibrotactile displays: A survey with a view on highly automated driving. *IEEE Transactions on Intelligent Transportation Systems*, 17(4), p. 897–907.

Pierce, R.S., Strybel, T.Z., & Vu, K.P.L. (2008). Comparing situation awareness measurement techniques in a low fidelity air traffic control simulation. *Proceedings of the 26th International Congress of the Aeronautical Sciences (ICAS)*. Anchorage, Alaska, USA.

Prince, C., Ellis, E., Brannick, M. T., & Salas, E. (2007). Measurement of team situation awareness in low experience level aviators. *The International Journal of Aviation Psychology*, 17, p. 41–57.

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), p. 31–43.

Recarte, M.A., Nunes, L.M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of experimental Psychology: Applied*, 9(2), p. 119–137

Recarte, M.A., Perez, E., Conchillo, A., & Nunes, L.M. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *The Spanish Journal of Psychology*, 11(2), p. 374–385.

Reimer, B. (2009). Impact of cognitive task complexity on drivers' visual tunneling. *Transportation Research Record: Journal of the Transportation Research Board*, 2138, p. 13–19.

Ries, F., Wolf, R., Olschowski, P., Dollken, M., & Demi, B. (2018). Disentangling Effects of Mental Workload and Visual Stimulus Properties on Pupil Size. Presentation of the Karlsruher Institut fur Technologie. DOI: 10.13140/RG.2.2.35819.85287
<https://www.researchgate.net/publication/323943103>

SAE (2018). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On Road Motor Vehicles: Surface Vehicle Recommended Practice. On-Road Automated Vehicles Standards Committee of the Society of Automotive Engineers International. Standard J3016.

Schwalm, M., Keinath, A., & Zimmer, H. D. (2008). Pupilometry as a method for measuring mental workload within a simulated driving task. In D. de Waard, F. O. Flemisch, B. Lorenz, H. Oberheid, and K. A. Brookhuis (Eds.). *Human Factors for Assistance and Automation*. Maastricht, the Netherlands: Shaker Publishing, p. 1–13.

Salmon, P.M., Stanton, N.A., Walker, G.H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring situation awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, 39, p. 490–500.

Senders, J.W. (1964). The human operator as a monitor and controller of multidegree of freedom systems. *IEEE Transactions on Human Factors in Electronics*, 1, p. 2–5. <https://doi.org/10.1109/THFE.1964.231647>

Senders, J. W. (1983). Visual sampling processes (Doctoral dissertation). Katholieke Hogeschool Tilburg.

Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., & Ward, J. L. (1967). The attentional demand of automobile driving. *Highway Research Record*, 195, p. 15–33.

Sivak, M. (1996). The information that drivers use: Is it indeed 90% visual? *Perception*, 25, 1081–1089.

Stanton, N.A., Salmon, P.M., & Walker, G. H. (2015). Let the reader decide: A paradigm shift for situation awareness in socio-technical systems. *Journal of Cognitive Engineering Decision Making*, 9, p. 44–50.

Stanton, N.A., & Young, M.S. (2000). A proposed psychological model of driving automation. *Theoretical Issues in Ergonomics Science*, 1(4): 315331. doi:10.1080/14639220052399131

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, 497–511.

Strybel, T.Z., Vu, K.P.L., Kraft, J., & Minakata, K. (2008). Assessing the situation awareness of pilots engaged in self spacing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52, p. 11–15.

Tatler, B.W. (2009). Current understanding of eye guidance. *Visual Cognition*, 17, p. 777–789.

Tatler, B.W., Hayhoe, M.M., Land, M.F., & Ballard, D.H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, 11(5): 5.

Tokuda, S., Obinata, G., Palmer, E., Chaparro, A. (2011). Estimation of mental workload using saccadic eye movements in a free-viewing task. *Engineering in Medicine and Biology Society, EMBC, Annual International Conference of the IEEE*, p. 4523–4529.

Tsai, Y., Viirre, E., Strychacz, C., Chase, B. & Jung, T. (2007). Task performance and eye activity: Predicting behavior relating to cognitive workload. *Aviation, Space, and Environmental Medicine*, 5, B176–B185.

Tsotsos, J.K. (1992). On the relative complexity of active vs. passive visual search. *International Journal of Computer Vision*, 7(2), p. 127–141.

Underwood, G., Crundall, D. & Chapman, P. (2011). Driving simulator validation with hazard perception. *Transportation Research Part F: Traffic Psychology and Behavior*, 14, p. 435–446.

Underwood, G. & Radach, R. (1998). Eye guidance and visual information processing: reading, visual search, picture perception and driving. In G. Underwood (Ed.) *Eye Guidance in Reading and Scene Perception*. Oxford, England: Elsevier, p. 1–28.

Van der Horst, R. (2004). Occlusion as a measure for visual workload: an overview of TNO occlusion research in car driving. *Applied Ergonomics*, 35, p. 189–196.

Victor, T.W., Harbluk, J.L., Engstrom, J.A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8, p. 167–190.

Wickens, C.D., Goh, J., Helleberg, J., Horrey, W.J., & Talleur, D.A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors*, 45(3), p. 360–380.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering Psychology and Human Performance* (Third edition). New Jersey, USA: Prentice Hall.

Yarbus, A.L. (1967). *Eye Movements and Vision*. New York, USA: Plenum Press.

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), p. 365–375.

Appendix A. Developed Driving Research Tools

A.1. Driving Scene Capture

To create visual driving recording stimulus materials (e.g., such as used in Chapters 3.1 and 3.2) there are many options with different associated costs, conveniences, and resultant fields of view. Mounts for filming equipment are a first underlying consideration. Some in-car camera mounts are affixed to windows and/or dashboards and can come at professional grade costs if needed. Here, a fast, simple, and effective design was implemented with a budget of only a few US dollars. Figure 3.2.A.1. shows my specific arrangement of three compact cameras affixed via standard $1/4$ " machine screws to 2" x 4" wooden support bars that were drilled to fit within the existing head reset mount interlock system. Thus, forward facing and periphery views were able to be captured by a single robust setup. Future studies could extend such an apparatus to hold various cameras at differently desired angles for multiple data collection purposes (e.g., in-vehicle occupant monitoring, capture of the driving scene from different passenger points of view, etc.).



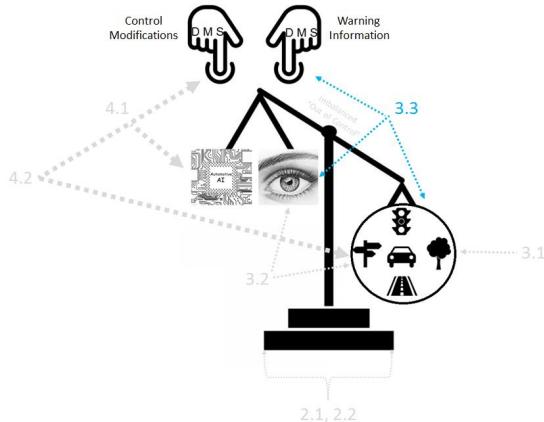
Figure 3.2.A.1. Low-cost driving scene video generation solution implemented to augment online videos collected and analyzed in the present chapters 3.1 and 3.2.

A.2. Automatic Clipping of Videos

After driving video recordings are collected, they may need to be parsed in various ways before being applied in an experimental research setting. For the studies in Chapter 3.1 and 3.2, it was desirable that different people would see not only different videos but across a set of exposures, comprehensive coverage of a single video could be enabled. Thus, a MATLAB function (included as copy/paste text below) was composed to automatically segment a longer duration video down into smaller clips of a specified length. Each video segment was set to start 1 second after the previous to ensure overlap within the dataset (i.e., for repeated measures reliability purposes). Note: this feature can be changed by adjusting the last number of ‘*while loop*’ iteration statement (e.g., 3rd to last line of code) to whatever the desired output video segment spacing might be (i.e., changing the iteration computation from ‘*i* = *i* + 1’ to ‘*i* = *i* + (*input_2*)’ would produce contiguous non-overlapping segments).

```
function vidChopperFun(input_1,input_2)
%Automatic segments of .mov video as overlapping clips of set size.
% input_1 = a video file, input_2 = desired segment size (secs)
% Each video segment starts 1 second after the previous.
% The first/last second of the full video is not included.
% Clip size must be at least 2 seconds less than full video length.
%Initialize limits
vidIn=VideoReader(input_1); input_2=input_2+1;
limit=round(vidIn.Duration,0)-(input_2);
i=1;
while i<=limit
    %Read in specific frames [start end]
    vidFrms=read(vidIn,[ (i*vidIn.FrameRate) ((i+input_2)*vidIn.FrameRate) ]);
    %Create a MATLAB movie struct from the video frames
    for k=1 : input_2*vidIn.FrameRate
        mov(k).cdata=vidFrms(:,:,:,:k); mov(k).colormap=[];
    end
    %Prepare and open the new file
    vidOut=VideoWriter(num2str(i)); vidOut.FrameRate=vidIn.FrameRate;
    open(vidOut);
    %Write each frame to the file
    for k=1 : input_2*vidIn.FrameRate
        writeVideo(vidOut,mov(k));
    end
    %Close the file
    close(vidOut);
    %Provide progress feedback
    disp(strcat('completed.',num2str(i),'.',num2str(input_2-1),...
        '-sec.clips.out.of.',num2str(limit),'.possible'));
    %Iterate to next whole second available from original video
    i=i+1;
end %End of while loop
end %End of function
```

Chap. 3.3) On-road Driver vs. Passenger Eye Eccentricity in a Conventional Car for In- vs. Out-of-the-loop “Drivenger” Monitoring in Automated Vehicles



In regards to the overall thesis big picture, this experiment serves to relate eye measurements with aspects of the on-road driving scenes/situational demands from which they were captured for the purposes of reducing the potential of driver monitor systems subjecting drivers/AV supervisors to unnecessary levels of over-alerting. Within an on-road study environment, Chap 3.3. investigates a different characterization of eye-scene relations than was able to be determined in the laboratory environment of Chap 3.2 (where scene demands could be more precisely measured and safely manipulated). Eye movements of drivers are contrasted with eye movements of passengers because while both are naturally in the same vehicle in the same driving environment, they are artificially divided in their responsibilities, and hence possess and represent different imposed attentional task demands (drivers being by definition in the control loop of driving, and passengers being by definition out of the control loop of driving). A continuous percentage distance eye eccentricity measure (ECC) discriminated at the level of momentary events, better at the level of individual participants and with longer measurement windows, and best when situated aspects such as vehicle speed and traffic count were also taken into account. Importantly, the eye eccentricity of all drivers safely rose (including prolonged periods of looking off-road) and fell across the driven trips where real-world driving scene task demands also naturally varied between relatively higher and lower demands.

Adapted from:

Cabral, C.D.D., Petrovych, V., de Winter, J.C.F., & Happee, R. (under review). On-road driver vs. passenger eye eccentricity in a conventional car for in- vs. out-of-the-loop “drivenger” monitoring in automated vehicles.

Abstract

Objective: To detect an out-of-the-loop driver state using eye-based criteria. **Background:** Many automated vehicles (AVs), whether released as SAE Level 2 or developing via on-road testing as SAE Level 3/4, require mentally '*in-the-loop*' human supervisors despite removing continuous hands/feet involvement. Driver monitor systems (DMS) can trigger when attention deviates from conventional in-control driver levels towards being more passenger-like (i.e., a '*drivenger*'). However, too many false alarms can undermine human trust, reliance, and acceptance of automatic alerts. **Methods:** Drivers and passengers simultaneously wore eye-tracking glasses on 32 on-road driving trips. An eye eccentricity (ECC) measure was computed as a mean percentage distance whenever eyes left a window-calibrated coordinate center gaze point. Impact of window size/levels and situated aspects (speed, steering angle, traffic count) on DMS alerting performance were assessed via ROC curves. **Results:** ECC was significantly higher for passengers. ECC discriminated between drivers and passengers both at the level of individual participants (based on the participant's average ECC score) and at the level of events (based on the momentary eccentricity score of an off-center looking event). ECC-based driver/passenger detection discrimination performance was improved by longer measurement window periods and consideration of vehicle speed (for windows shorter than 1 minute) and traffic count with vehicle speed (for windows longer than 1 minute). **Conclusion:** Our introduced measure differentiates in- vs. out-of-control eyes. For DMS, we recommend use of relative moving window averages and situated criteria to reduce false alarms. **Application:** Passengers in conventional vehicles can help refine measures for AV driver vigilance.

1. Introduction

1.1. Background

Classic driver distraction problems suggest a turn towards driving automation to serve as an impactful safety solution. Crash data from 2010 showed that 17 percent (an estimated 899,000 crashes) of all police-reported crashes in the U.S. involved some type of driver distraction (NHTSA, 2013a). In a 50 year review of driving safety research, Lee (2008) relates that crashes are often caused by drivers failing to look '*at the right thing at the right time*' and cites evidence suggesting that even short glances away increase crash risk (Klauer et al., 2006). Meanwhile, automated vehicles (AV) are recently emerging in terms of on-market automotive features (Mays, 2018) as well as on-road tests and developments (CA DMV, 2018). AV technology is often motivated by safety claims to address human errors (e.g., NHTSA, 2017) such as the above distraction issues. For example, NHTSA (2008) (where aberrant driver states and behaviors were found associated in a majority of fatal crashes) is frequently cited attributing 90% or more of causal blame towards the human rather than the vehicle or the environment. However, vehicles that range between being able to do *some* or *almost all* of the driving inherently lack a *full* authority and thus require human oversight and back-up. Recent tests from Euro NCAP (2018) concluded that '*cars, even those with advanced driver assistance systems, need a vigilant, attentive driver behind the wheel at all times*'. Likewise, AAA (2018) expressed a cautionary sentiment towards consumers becoming disengaged during partially automated driving. Inadequate safety-driver supervision was implicated in the first widely reported pedestrian fatality of an autonomous vehicle (Coppola & Frank, 2018).

1.2. Automation-induced 'out-of-the-loop' concerns

New kinds of inattention issues may arise when humans are required to monitor driving automation. Historically, a wide body of human factors research has suggested expectations for problems in placing people in this sort of role. Endsley and Jones (2012, Chapter 10) summarize hindrances to situation awareness while supervising automation due to issues of complacency, passive processing of information, and quality of system feedback. Concerns surrounding limited human vigilance in supervising automated processes can be traced back to Mackworth (1948) and have been observed to exist for both simple and more complex kinds of monitoring tasks (Parasuraman, 1987). Across several studies, the situation awareness of air traffic controllers has been commonly observed to suffer when only monitoring rather than actively controlling aircraft (Endsley et al., 1997; Endsley & Rodgers, 1998; Metzger & Parasuraman, 2001). Adaptive automation concepts have been shown to effectively close feedback loops towards enhanced operator engagement (Parasuraman et al., 1996). In practice, a range of adaptively triggered functional outcomes can vary between differently designed alert-notification-warnings and/or transitions of control.

Recent studies suggest empirical evidence of a human deficiency in monitoring specifically for driving automation (Greenlee et al., 2018; Banks et al., 2018) while others contain eye tracking of AV drivers (Merat et al., 2014; Louw et al., 2016; Louw & Merat, 2017; Pampel et al., 2018) and so represent an apparent interest towards the topic of driver visual vigilance within AVs. By removing traditional in-the-loop motor control activities of hands and feet, and especially where remaining engagement is characterized as for exceptional rather than nominal circumstances, AVs might paradoxically prime drivers towards familiar passive passenger levels of attention akin to '*along-for-the-ride*' responsibility even if some may technically require full and active alertness. Thus, a collective interest is observed centered around the topic of catching and protecting against

'drivenger' states (i.e., where traditional levels of driving visual control might stray from active conventional drivers towards more passive passengers) as a newly introduced susceptibility resultant from what Banks & Stanton (2017) have dubbed the *'Driver Not Driving'* role in the automotive domain (e.g., an analog to the *'Pilot Not Flying'* role in aviation).

1.3. Driver Monitoring Systems (DMS) for both conventional vehicles and AVs

Monitoring activity by drivers is inherent across multiple levels of conventional driving as well as driving automation: for operational functions of lateral/longitudinal movements, for tactical functions of object/event detection, and for strategic functions of navigation (see Merat et al., 2018, esp. Fig. 2). Thus, meta-monitoring from driver monitoring systems (DMS) is expected to be valuable on account of the prevalence of what Merat et al. (2018) refer to as *'in/on-the-loop'* activity. DMS may utilize various behavioral, subjective, and physiological measures (see Dong et al., 2011, for a review). Behavioral measures such as steering movements are popular in on-market systems, but from a perspective of active safety, might be regarded as relatively *'reactive'* rather than *'proactive'* by their measurement of consequences rather than predictive indices. Subjective measures can be difficult to incorporate in real-time DMS and carry risks of inaccurate introspection—Schmidt et al. (2009) found a lack of ability in the self-assessment of vigilance after continuous monotonous driving. Physiological measures vary along a dimension of equipment obtrusiveness such as between electrodes (EEG, ECG, skin conductance), pressure transducers (respiratory responses) and cameras (eye, face, and body tracking). Previously, state-of-the-art releases of DMS for inattention while supervising driving automation had thus far in the majority relied on steering measures indicative of hand placement (e.g., Tesla's *'Autopilot'*, Volvo's *'Pilot Assist II'*, Audi's *'Adaptive Cruise Assist'*, BMW's *'Active Driving Assistant Plus'*, Daimler's *'Distronic Plus'*). Recently, however, gaze/head based camera DMS are now beginning to reach the AV functionality market as well (e.g., GM's *'Driver Attention System'*, Subaru's *'Driver Focus'*, Audi's *'zFAS'*) ahead of reports (e.g., Yoshida, 2018) of increased demands and roadmap releases from Euro NCAP targeting (presumably camera-based) DMS as a primary safety standard by 2020.

In particular, eye-tracking technology in DMS for inattention in supervising driving automation is expected to show promise for many reasons. It maintains face validity benefits where overt fixations are generally assumed to indicate attention in the sense of information uptake (e.g., Just & Carpenter, 1980; Shoaeeizadeh, et al., 2016). In accordance with a model of directional relations between cognitive constructs (i.e., Heikoop et al., 2015), a state of attention is expected to occur later in a chain of related states and is thus a more preferable measurement construct than earlier states that it presumably subsumes (i.e., fatigue and/or workload). Furthermore, driving performance decrements from distraction appear more capable of being resolved in comparison to recovering from fatigue (Hancock, 2013). Lastly, a wide body of research has previously established: a pre-dominant importance of visual information for driving (Sivak, 1996), the expected frequencies with which drivers look to specific objects (Gordon, 1966; Serafin, 1994; Green, 2001), scene-situated variations in gaze such as with route familiarity (Mourant & Rockwell, 1970), curvy roadways (Land & Lee, 1994), car following (Tijerina et al., 2004), overtaking (Gray & Regan, 2005), intersection negotiation (Romoser, 2013), and general increased cognitive task complexity (Reimer, 2009), the modeling of driver visual sampling (Senders et al., 1967; Salvucci & Gray, 2004) and distraction behavior (Sheridan, 2004; Liang et al., 2012), as well as multiple eye-based DMS developmental applications (Dinges et al., 1998; Smith et al., 2000; Ji & Yang, 2001; Ohn-Bar & Trivedi, 2016).

1.4. Situating eye measures of aberrant driver attention

A conceptual complication and conflict for the eye-tracking of drivers is suggested from the top-most fundamental principle from U.S. federal guidelines regarding driver distraction (NHTSA, 2013b) where it is seen that not all looking around behavior is necessarily bad: *'the driver's eyes should usually be looking at the road ahead'* (emphasis added). Klauer et al. (2006) seem to agree from their own conclusions that *'short, brief glances away from the forward roadway for the purpose of scanning the driving environment are safe and actually decrease near-crash/crash risk'*. Ultimately, however, the challenge of ambiguity in good vs. bad driver visual behavior is downplayed by Klauer et al. (2006) by their proposition of a hard and fast rule they relate as *'glances totaling more than 2 seconds for any purpose increase near-crash/crash risk by at least two times that of normal, baseline driving'* (emphasis added). Hence, such a 2 second rule (cp. Rockwell, 1988) has been since adopted in the NHTSA guidelines regarding the amount of time that the driver's eyes are drawn away from the roadway during the performance of a task (NHTSA, 2013b).

An absolute and fixed criterion based only on timing seems inconsistent with a variety of research practices and foci around different scene-dependent factors in driving: day/night, straight/curved roads, young/old drivers, familiarity/novelty, presence/absence of lead car and/or other traffic, etc. that can be observed across a 6 decade span of research regarding how long and where drivers look around (Table 3.3.1). Victor et al. (2005) lament that *'It seems unnecessarily restrictive that evidence of a single glance longer than two seconds by a single subject could create a fail situation'*. Furthermore, visual occlusion techniques have shown both durations longer than 2 seconds 'away' and many other situation dependencies. Averaged voluntary occlusion periods evidenced in Godthelp et al. (1984) ranged from 2.5 to 5.5 seconds where it was concluded that drivers use a relative basis of time available (i.e., including aspects of lane position and vehicle velocity) to determine their visual information needs rather than some constant amount of time. Furthermore, Victor et al. (2005) relate the classic findings of Senders et al. (1967) as *'drivers dramatically increase eyes-off-road-times as speed is reduced. This result indicates that glance duration as a measure must be considered in relation to the driving demand imposed by the situation, for example speed.'* Extensions from classical control theory (error/uncertainty nullification) perspectives posit prominence of mental models to decide on contextualized probabilities/expectancies and effort (Sheridan, 2004) to serve information bandwidth models such as theorized from the likes of Senders et al. (1967), Wierwille (1993), Mourant & Ge (1997) and Courage et al. (2000) to drive periodic visual sampling in automobile control. Kircher & Ahlstrom (2017) introduced a theory of minimum required attention (MiRA) that accounts for adaptive human visual behavior where *'a driver is considered attentive when sampling sufficient information to meet the demands of the system'*.

Table 3.3.1. Aspects of measuring where and for how long drivers look around while driving to inform definitions of nominal and off-nominal looking. *Sub-set concerning daytime straight road driving.

Year	First Author, Last Name	Looking Time Definition Aspects	Looking Distance Definition Aspects
1967	Senders	300 second window periods: self-chosen occlusion intervals	Discrete; Binary; Visual occlusion device in front of the eyes. Open or shut.
1975*	Rackoff	30 second window periods	Discrete; AOI set; (6)

Year	First Author, Last Name	Looking Time Definition Aspects	Looking Distance Definition Aspects
1977*	Shinar	Unspecified window period: two straight segments among 22 curves across 34 km	(6 exterior, 0 interior, 0 other); Ahead = 3 by 6 degrees around the focus of expansion; Manual annotation
1984	Godthelp	450 second window periods: self-chosen occlusion intervals	Discrete; AOI set; (4) (2 exterior, 2 interior, 0 other), Ahead = 1.6 degrees to the right and 0.7 degrees above the focus of expansion; Manual annotation
1989*	Olson	30 second window periods: based on total run distance of 1 mile long for 120 seconds and reported 0.25 mile of the straight segment	Discrete; Binary; Visual occlusion device in front of the eyes. Open or shut.
1994*	Serafin	50 second window periods: based on posted speed limit of 50 mph, and distance of about 0.7 mile of the straight segment	Discrete; AOI set; (8) (6 exterior, 1 interior, 1 other), Ahead = 3 different possible AOIs: center of road, lead car, far field; Manual annotation
2000	Recarte	30 second window periods	Discrete; AOI set; (15) (9 exterior; 4 interior; 2 other), Ahead = 2 different possible AOIs: right lane, far field; Software annotation
2005	Victor	30 second window periods Percentage of gaze samples in a defined road center area	Continuous; Standard deviation of gaze position (angle) relative to focus of expansion; Software annotation
2006	Zhang	60 second window periods	Discrete; Binary; AOI set; (2) (0 exterior; 0 interior; 2 other), Ahead = a circle of 16 degrees surrounding a modal (most frequent) gaze angle position; Software annotation
2007	Donmez	3 second window period: degree of distraction as a function of current off-road glance duration	Discrete; AOI set; (4) (1 exterior; 3 interior; 1 other), Ahead = ; both vertical and horizontal gaze angles were between +12 and -12 degrees at the focus of expansion on the horizon line; Software annotation
			Discrete; Binary; AOI set; (2) (1 exterior; 1 interior; 0 other), Ahead = ; 'at the road';

Year	First Author, Last Name	Looking Time Definition Aspects	Looking Distance Definition Aspects
		compared to total off-road glance duration during the last 3 seconds	Software annotation
2008	Reyes	60 second window periods	Continuous; Standard deviation of gaze position (angle) relative to focus of expansion; Software annotation
			Discrete; Binary; AOI set; (2) (2 exterior; 0 interior; 0 other)
			Ahead = less than 5 degrees either direction of mode horizontal fixation Right side = greater than or equal to 5 degrees to the right of the mode horizontal fixation; Software annotation
2008	Zhang	Variable up to 30 second window period: The time window was re-configurable within the range of 1-30 secs, i.e., implemented a 4.3 second window	Discrete; Binary; AOI set; (2) (1 exterior; 1 interior; 0 other)
			Ahead = +/- 24 degree horizontal, +/- 24 degree vertical, rectangular forward area; Software annotation
2009	Kircher	2 second window period: 2 sec time buffer starts to deplete upon moving gaze away from Ahead; 0.1 sec latency in returning gaze Ahead before refilling the time buffer; 1 sec latency in moving gaze away from Ahead but to speedometer or a mirror	Discrete; AOI set; (6) (2 exterior; 4 interior; 0 other)
			Ahead = +/- 45 degrees horizontal, +45/-22.5 degrees vertical; Software annotation
2010	Weller	1 second window periods based from subsections of 25 meters (i.e., 90 km/h)	Continuous; Standard deviation of gaze position (pixel distances); Software annotation
2014	Merat	10 second window period: If driver looked away from 'road centre' for 10 secs or more.	Discrete; Binary; AOI set; (2) (0 exterior; 0 interior; 2 other), Ahead = ellipse with a 10 degree major and 6 degree minor radius; Software annotation
		5 and 60 second window periods: PRC data were plotted at 5 sec intervals, for the first 60 secs after disengagement of driving controllers.	Discrete; Binary; AOI set; (2) (0 exterior; 0 interior; 2 other), Percentage of gaze samples in a defined road center area, Ahead = a circle of 6 degrees surrounding a modal (most frequent) gaze angle position; Software annotation

Year	First Author, Last Name	Looking Time Definition Aspects	Looking Distance Definition Aspects
2015	Vicente	0.033 second (30hz) instantaneous as well as 10 second window evaluation period: If the intersection point lies outside of the defined on-the-road area, an alarm is triggered; as well as a percentage of frames correctly predicted in terms of eyes on/off-the-road during ten second periods	Discrete; Binary; AOI set; (2) (1 exterior; 1 interior; 0 other), Ahead = windshield plane; Software annotation
2016	Louw	1 second time window periods across 3 consecutive seconds; Percentage of gaze samples in a defined road center area	Discrete; AOI set; (5) (5 exterior; 0 interior; 0 other), Ahead = a circle of 6 degrees surrounding a modal (most frequent) gaze angle position; Software annotation
2017	Louw	Varying time window periods: 100 seconds, 30 seconds, 8 seconds, 3 seconds	Continuous; Standard deviation of horizontal and vertical gaze positions (angle); Software annotation
2018	Pampel	5 and 60 second window periods: by splitting the one-minute period into 5.0-second time bins; From the previous 10 seconds to the future 10 seconds Percentage of gaze samples in a defined road center area	Discrete; AOI set; (2) (1 exterior; 1 interior; 0 other), Ahead = Within 20 degrees horizontal and 15 degrees vertical around the mean fixation point; Software annotation Continuous; Standard deviation of horizontal gaze positions (angle); Software annotation

Specific nuances of measurement and application still remain an open point of research for DMS algorithms and systems integration alerting criteria. While Klauer et al. (2006) used manual reductionist methods (i.e., human annotators), a real-time DMS would require an a-priori computerized definition of what constitutes '*looking ahead*' vs. '*glancing away*' in terms of boundary definitions in both time and space. As seen in Table 3.3.1, components of time and distance in defining driver distraction present relatively more variability than definitive consensus in how to directly proceed with functional criteria for building a real-time DMS system. Timing measurement aspects vary in window size and number (i.e., sub-windows) while distance measurement aspects vary by nature of being continuous or discrete (and if discrete, then in the number of defined boundaries). In general, a popular approach appears to typically discretize pre-defined areas of interest – AOIs (e.g., percentage road center – PRC, AttendD, etc.).

1.5. A new distance-based measure of eye eccentricity (ECC)

In principle, the previous measures are distance agnostic and bound to pre-defined central focus area boundaries. However, previous research suggests that eyes of drivers might frequently change their points of reference—with different sized cars and turn radiiuses (Olson, 1964); in the presence of a lead vehicle (Mourant et al., 1969), and situational demands of signage, other vehicles, and road edge markings (Mourant & Rockwell, 1970); while approaching and transiting curves (Laya, 1992; Land & Lee, 1994), with varying levels of experience across rural, suburban, and expressway roads (Crundall & Underwood, 1998); and between near and (differing) far points (Salvucci & Gray, 2004); as well as to scan a number of off-center driving-related areas/objects (e.g., an assumption of Kircher and Ahlstrom, 2009). While AOI-based measures can provide foundational detections of overt distraction (e.g., looking too long at a secondary task on an in-vehicle display or mobile device), they are challenged to account for more covert inattention issues such as '*looked-but-failed-to-see*' (Herslund & Jorgensen, 2003) errors where driver eyes can fall within the normative AOI bounds (and/or for the normative durations) but with a disconnect to perceptual/cognitive processes.

Eye movement measures can also be defined in ways that do not require labeled AOI boundaries which can be complicated by the diversity of vehicle interiors, driving scene exteriors, and growing/shrinking AOIs from near/far 3D movements. Thus, AOI-based driver eye behavior measures stand to be complemented and extended with applications that incorporate continuous gaze location/extent (e.g., standard deviation of gaze - SDG) as another dimension of eye-movement resolution. For example, Louw & Merat (2017) used a remote eye tracker mounted on the dashboard of a driving simulator and found an increase in horizontal gaze dispersion via SDG measures for conditions of automated vs. manual driving. Like SDG, an eye eccentricity – ECC measure (i.e., from a head-mounted eye tracker without a 3D world model) might be defined and used to capture continuous distances of eye movements beyond discrete bounded thresholds. Unlike conventional static AOI methods, ECC can make use of a central tendency coordinate point (average gaze location over a measurement/period of interest) from which to relatively compute a dynamically calibrated off-center distance. In sum, a new ECC measure could extend previous eye measures by relatively moving around with and as the head and eyes move around, with less restrictive absolute definitions. Thus, ECC could be a kind of DMS measure that can differentiate '*in-the-loop*' eyes that are up, on the road, and moving around versus '*out-of-the-loop*' eyes that are also up, on the road, and moving around but in different manner (i.e., different distances).

1.6. Present multi-phase study motivations and aims

1.6.1 Phase 1: Replication of driver-passenger eye differences and validation of ECC

A multi-algorithmic study of Liang et al. (2012), hypothesized to obtain an eye-distance-based effect in crash-risk prediction performance (from a naturalistic dataset), but ultimately attributed a lack of obtained differences to their estimated location data from manually coded video data, while suggesting further investigations with more precise measures of visual angle derived from eye-tracking data. Using electrooculogram (EOG) techniques (i.e., electrical potential measurements from near-eye electrodes), Takeda et al. (2016) found a difference between in-the-loop drivers and out-of-the-loop passengers in terms of the number of small/large sized saccadic eye movements (i.e., some distance based differences).

In Phase 1, we aimed to define and validate a new inattention construct via on-road measurements to differentiate '*in-the-loop*' eye movements versus '*out-of-the-loop*' eye movements. Analyses from the drivers versus passengers study of Takeda et al. (2016) were replicated to ensure similar attentional differences existed between our own set of drivers and passengers before applying our new measure of off-center eye eccentricity (ECC). A preliminary analysis of the present data is provided by Cabrall et al. (2017).

1.6.1 Phase 2: Hypothetical application of DMS using ECC

As an alerting agent, a DMS bears a burden beyond its objective detection performance towards establishing credibility with the driver. Trust is a major component for effective human-automation interaction (e.g., Lee & See, 2004) and its constituent components of reliance and compliance have been identified by Parasuraman and Wickens (2008) to be determined by the thresholds that designers use to balance automation misses and false alarms. The negative subjective experience of over-alerting has been commonly referred to as a '*cry-wolf*' effect that diminishes trust and detracts from warning compliance. Automation mistakes on tasks deemed easy for humans are particularly detrimental to trust development (Madhaven et al., 2006). Thus, counterproductive effects could be expected if a driver is automatically assessed as being inattentive when they believe otherwise (e.g., while looking away from the road while at a red light). Such considerations suggest a purely human-centric DMS (i.e., that responds only to physiological/behavioral measures of a person) could be less useful than a situated DMS (i.e., that also accounts for aspects of the situation the person presently resides within).

In Phase 2, the aim was to explore implementation criteria of a hypothetical DMS to reduce perceived false alarms of that DMS. For the second aim, effects of using ECC under different threshold levels of eccentric events and aggregation levels (at the level of individual events vs. at the level of individual drivers) and of accounting for common (automated) vehicle telemetry items (i.e., vehicle speed, steering angle, presence of lead vehicles) were all examined, with a focus on the trade-off between misses and false alarms.

2. Methods

2.1 Participants

The experiment was completed in November 2016 by 16 pairs of participants (78% male, 22% female, mean age = 27.3, SD age = 2.4) recruited from the Delft University of Technology. Written informed consent was obtained under the approval of a Human Research Ethics Committee (On Road In Vehicle Eye Tracking: Drivers and Passengers, 26 September 2016). Each participant had normal or corrected-to-normal vision and reported having obtained their initial driver's license for at least more than one year prior to the experiment. Participant pairs were formed around a quasi-experimental variable of familiarity, such that half of the pairs knew one another well, whereas with the other half, participants were not known to one another in advance. analyses pertaining to this aspect, however, remain to be pursued in future follow-on studies.

2.2 Driving route and procedures

The driving route began from Leeghwaterstraat 21 and proceeded across campus via Jaffalaan and Mekelweg/Christiaan Huygensweg, then continued southbound on Schoemakerstraat, westward

along Kruithuisweg, joined the A4 highway northbound until exit 12 for route N211, at which point the route crossed over the highway and returned along the same roads in reverse direction (Figure 3.3.1). The route was selected to contain a wide variety of driving situations (e.g., road geometry, traffic, signage). The full route was completed as one trip of about 20.0 km and around 30 minutes on average, and repeated per pair with a switching of driver/passenger role, for a total of 32 trips.



Figure 3.3.1. The route covered mixed urban and highway roads. Representative screenshots are provided for various route segments at number-labeled points on the map.

Drivers were given no instructions other than to drive as they normally would (i.e., in a safe manner). Passengers also began without any instructions to deviate from their normal behaviors, but at the turnaround point they were given a piece of paper to covertly assign an experimental manipulation: '*Please imagine that you are doing the driving. So try to pay attention and behave with your eyes as if you are currently driving. You do not need to move your hands/feet like a driver*'. In lieu of naturalistic observation motivations for any role, no restrictions were expressed per conversation, use of electronic devices, etc.

2.3 Apparatus and measurements

Both passenger and driver participants wore UV shielded eye-tracking glasses from SensoMotoric Instruments (SMI) coupled by a single USB cable each to their own dedicated Samsung Galaxy smartphone running only the eye-tracking software and were held by ride-along experimenters in the backseats (Figure 3.3.2). The car driven was a 2014 Toyota Prius Hybrid passenger vehicle with automatic transmission without use of any cruise control and was equipped with driving research telemetry for vehicle state and control input data.



Figure 3.3.2. Passenger and driver wearing minimally invasive eye-tracking glasses.

The glasses recorded eye measurement samples (60 Hz), with gaze data indexed by a 960 x 780-pixel coordinate grid in respects of a viewing plane of the forward facing camera above the nose bridge of the glasses. The eye-tracking software used virtual geometrical dimensions of the viewing plane to automatically compute and log its gaze coordinates as if on top of such a screen: 960 mm (wide) x 780 mm (tall) with a depth location of 145 mm (in-front). Missing data (e.g., due to blinks) or data points out of the screen bounds were removed and subsequently linearly interpolated.

Driver and passenger ECC scores were computed based on the distance to a calibrated center region:

- (1) The median x and y gaze coordinates of a measurement/analysis period were subtracted from the original x and y gaze coordinates, to obtain a calibrated value of the gaze coordinates around (0,0).
- (2) The Euclidean distance was computed from each gaze sample location to the central coordinate point (0,0), and divided by 600 (and multiplied by 100) to result in a percentage of distance from the center.
- (3) The eccentricity score for a particular period of interest is the mean of the distance scores in a selected period of interest.

2.4 Analyses

For the first phase analyses, three periods of interest were applied that varied in duration and expected driving situational demands.

- (1a)/(1b) "Post/Pre Task", each about 100 seconds. The first period revolved around a transition from a naturalistic passenger to an (enacted) driver role as the portion of time for the few minutes immediately before compared against the time period immediately after the passenger task instruction presentation (with 20 seconds before/after the task start removed to exclude reading/processing of task instructions). For the passengers, '*post-task*' data are regarded as '*pseudo drivers*' attempting to represent visual control whereas the '*pre-task*' data of the passengers are regarded still as natural (untasked) freely varying passenger eye data.
- (2) "Entering A4", about 45 seconds. The second period involved entering and merging onto a highway. The first highway on-ramp and merging period where it was assumed a driver would be likely to prioritize and evidence high levels of dedicated driving control visual behavior. Both driver and passenger eye data are included.
- (3) "Gate to Task", about 950 seconds. The third period was an extended period to capture the entire first half of the drive, from the start of the trip (leaving the parking lot gate) up until the start of the passenger task manipulation.

Because eye-tracking data are susceptible to missing values which might affect data validity, we removed participant pairs if more than 20% of data were missing or out of the forward-facing screen bounds for either driver or passenger.

3. Results

Equipment errors resulted in complete eye tracking data loss from 2 drivers and 2 passengers. Furthermore, for 2 drivers and 2 passengers, more than 20% of the gaze data were missing. Thus, eye-tracking data were available for 28 of 32 drivers, and for 28 of 32 passengers.

A distribution of the eccentricity values showed clear differences between drivers and passengers (Figure 3.3.3). Drivers were more likely to look ahead (< 13%), while passengers were more likely to look away from their central point (> 13%).

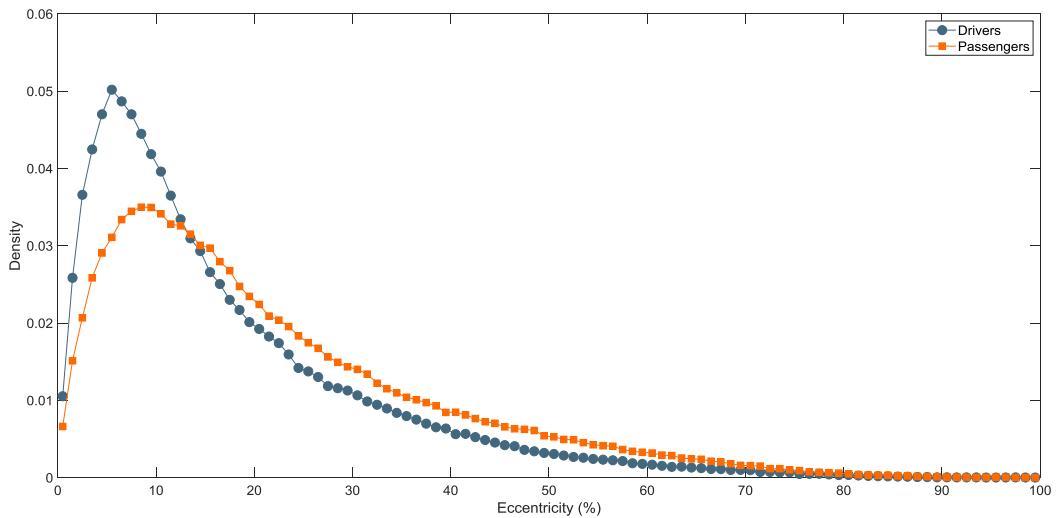


Figure 3.3.3. Distribution of eccentricity values at the sample level (60 Hz) for the “gate to task” period (about 950 s of driving per trip). The distribution was calculated for each driver and passenger separately, and subsequently averaged across the drivers/passengers. The eccentricity values were divided into 100 one-percent bins. The two density curves have been normalized so that the sum of the 100 data points equals 1.

Mean eccentricity scores were computed for drivers and passengers (Figure 3.3.4). Independent-samples t -tests showed that passengers had statistically higher eccentricity scores as compared to drivers for three of the four conditions shown in Figure 3.3.4, $t(54) = 4.59, p < 0.001$, $t(54) = 3.55, p < 0.001$, $t(54) = 4.78, p < 0.001$, $t(54) = 1.59, p = 0.118$, respectively. Thus, passengers exhibited significantly higher eccentricity scores than drivers, except in the post-task period where passengers had the task to look as if they were a driver.

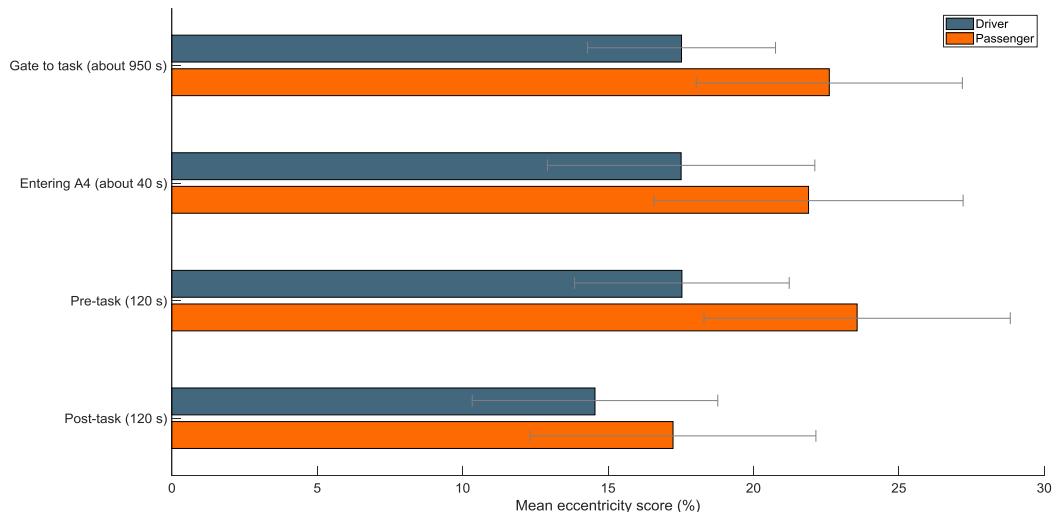


Figure 3.3.4. Average eccentricity scores for drivers and passengers. Error bars run from the mean \pm 1 standard deviation. The means and standard deviations were calculated for 28 drivers and 28 passengers.

The gray line in Figure 3.3.5 is the ROC curve for the eccentricity scores at the level of the drivers ($n = 28$) and passengers ($n = 28$) for the gate to task period. It can be seen that reasonable discrimination is achieved, e.g., for a hit rate of 82% (i.e. detecting that the passenger is indeed a passenger), there is a false alarm rate of 36% (i.e., falsely detecting that the driver is a passenger). Figure 3.3.5 also shows the ROC curve at the sample level in green. More specifically, we calculated the true positive rate versus false positive rate for all individual eccentricity samples of the experiment ($n = 1,611,825$ for drivers, $n = 1,611,383$ for passengers). It can be seen that discrimination between driver and passenger at the sample level is less strong than at the participant level. This means that there is poor discrimination between drivers and passengers based on the eccentricity of a single sampling instance. Figure 3.3.5 also shows that the discrimination between drivers and passenger becomes better when applying a moving average on the eccentricity scores. That is, when aggregating eccentricity data for a minute or five minutes, it becomes reasonably possible to distinguish drivers from passengers.

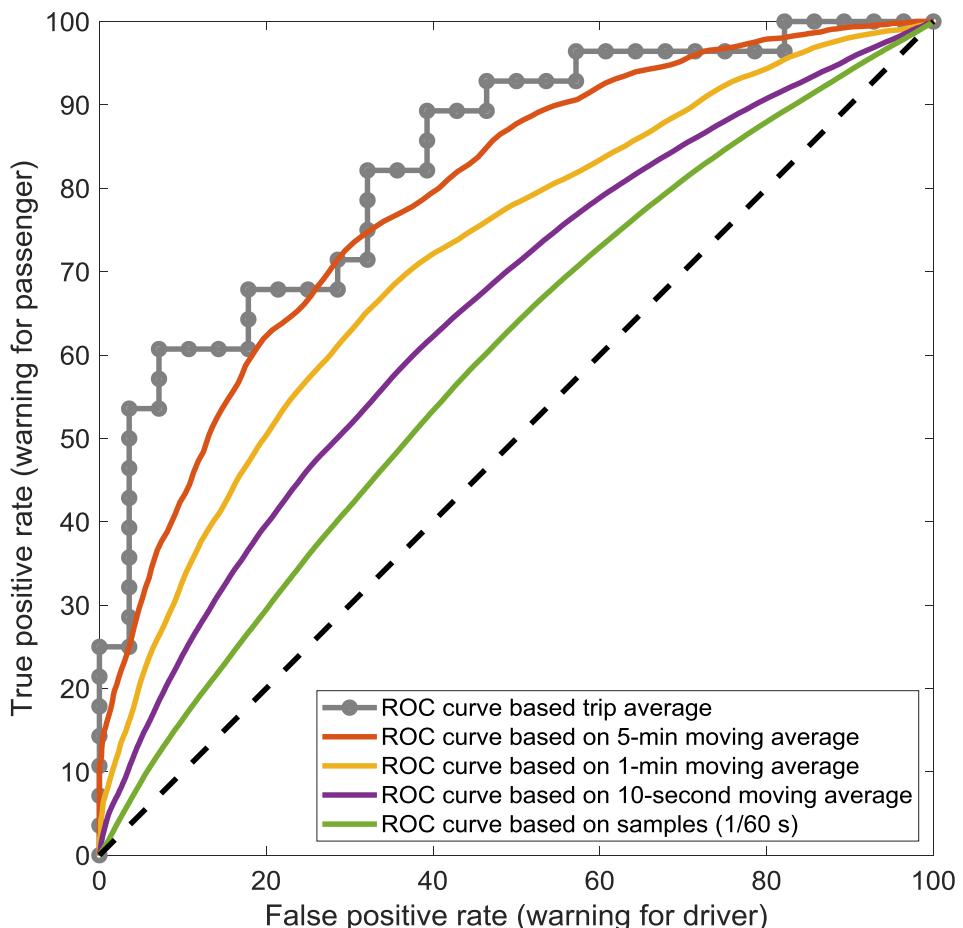


Figure 3.3.5. Receiver operating characteristic (ROC) curves for a hypothetical driver monitoring system which issues a warning when eccentricity exceeds a threshold level. The ROC curve is provided at the level of measurement samples and at the level of trips. The figure is based on the gate to task period.

One explanation for the poor discrimination for individual samples is illustrated using Figures 3.3.6–9. In these figures, evidence is suggested that shows that the eccentricity level was situation-dependent. For a first example, when taking a turn (recognizable by low speed and a high steering angle) with elevated potential for interactions with other traffic, eccentricity scores of drivers and passengers were high (i.e., first inset of each Figures 3.3.6–9). As a second example, in a situation when driving scene demands were more predictable/stable (recognizable by high speed, low steering angle, and low traffic count), eccentricity scores of drivers also rose several times to overlap with passenger levels (i.e., second inset of each Figures 3.3.6–9). These findings suggest that a DMS should be context-dependent, by taking into account the viewing demands of the situation.

A possible solution for improving the accuracy of the classifier is to discard (e.g., refrain from potentially over-alerting during) situations where off-center looking might be expected to increase as natural/safe adaptation to relative extremities of high/low (visual) driving scene task demands. Figure 3.3.10 illustrates that classification becomes better when excluding moments where the vehicle was driving slower than 20 km/h, when the traffic count was less than or equal to 1, and best overall when accounting for both sources of information. In particular, for moving window average sizes of around 1 minute or less, speed constraints had the largest benefit to discrimination performance while for larger window sizes, traffic count constraints provided additional benefits in conjunction with speed constraints.

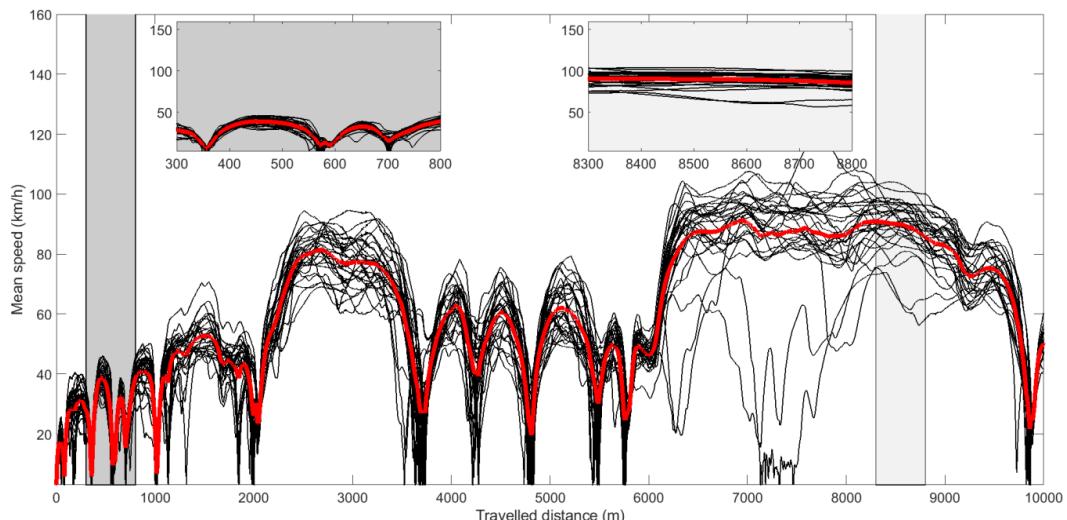


Figure 3.3.6. Vehicle speed as a function of travelled distance for 30 trips (2 trips were excluded because the driver took a wrong turn and so the total travelled distance different from the others). The black lines represent the speed of the 30 individual trips, whereas the red line represents the average of the 30 trips. The figure is based on the gate to task period.

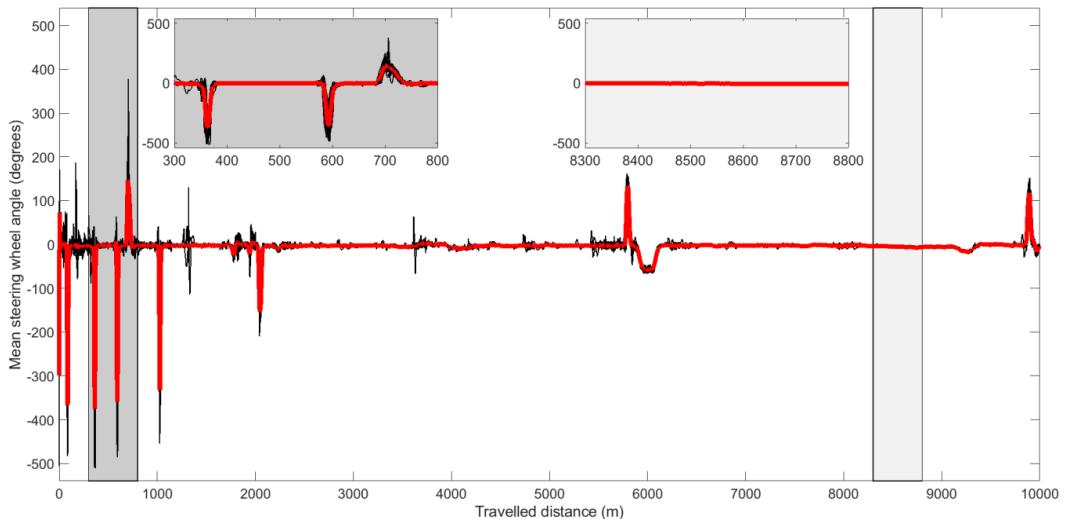


Figure 3.3.7. Steering wheel angle (left = positive, right = negative) as a function of travelled distance for 30 trips (2 trips were excluded because the driver took a wrong turn and so the total travelled distance different from the others). The black lines represent the steering angles of the 30 individual trips, whereas the red line represents the average of the 30 trips. The figure is based on the gate to task period.

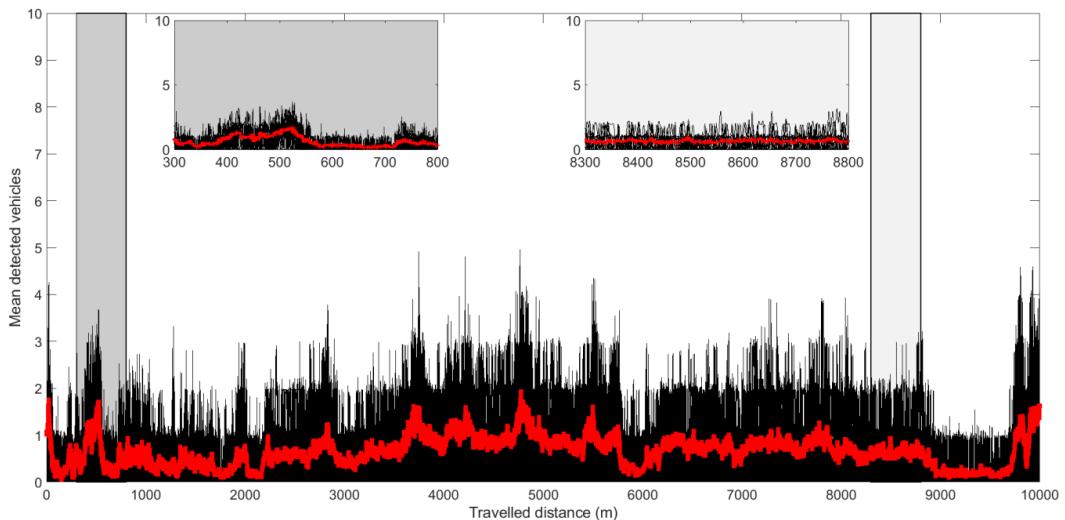


Figure 3.3.8. Approximated traffic count as a function of travelled distance for 30 trips (2 trips were excluded because the driver took a wrong turn and so the total travelled distance differed from the others). The black lines represent the approximated traffic count of the 30 individual trips, whereas the red line represents the average of the 30 trips. The figure is based on the gate to task period.

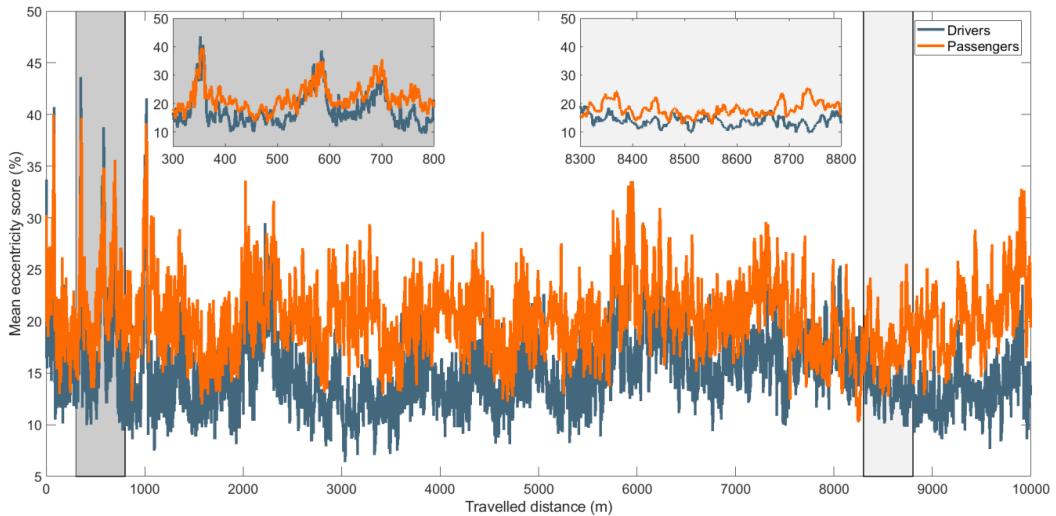


Figure 3.3.9. Mean eccentricity scores as a function of travelled distance of 27 drivers and 26 passengers (2 trips were excluded because the driver took a wrong turn and so the total travelled distance different from the others). The figure is based on the gate to task period.

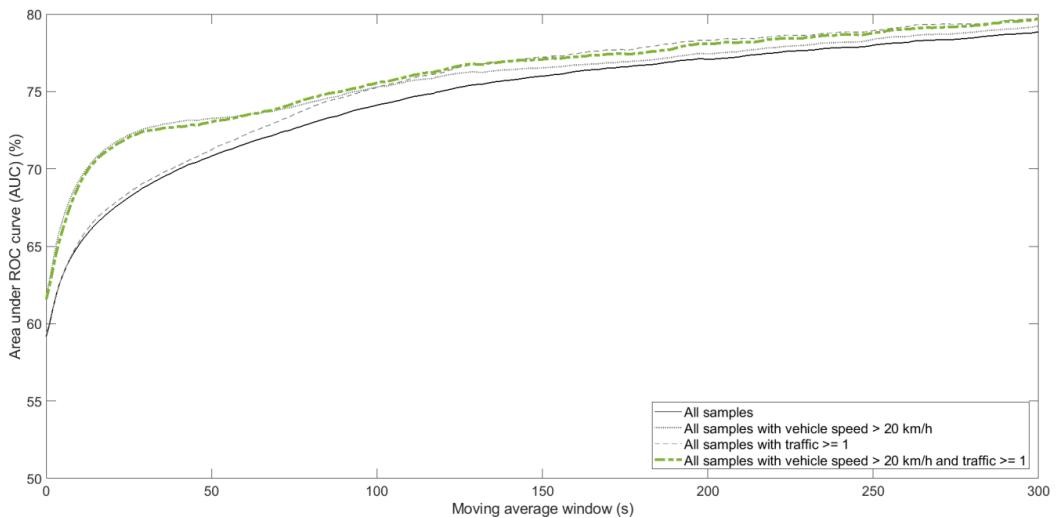


Figure 3.3.10. Area under the ROC curve (%) for different moving time windows of the eccentricity values for all samples (corresponding to the ROC curves in Figure 3.3.5) and for all samples for which the vehicle drove faster than 20 km/h and 1 or more vehicles were determined to be present (from automatic visual detection) in the driving scene.

4. Discussion

In Phase 1, our aim was similar to Takeda et al. (2016) of detecting possible ‘drivenger’ like lapses attention of supervisors of future AVs via eye measurement differences between in-control drivers and not-in-control passengers in a conventional vehicle. We replicated results of Takeda et al. (2016) in finding passengers (as compared to drivers) to exhibit a higher variance of gaze as significantly higher levels of ECC were found in passengers compared to drivers. With more precise

measurement of visual orientation (e.g., increased resolution of automatic continuous coordinates rather than manual annotations of inside/outside of AOIs) we obtained an effect of distance consistent with the (unattained) hypothesis from Liang et al. 2012, i.e., greater visual distances associated with less driving control. On account of safety (and face validity) concerns we did not attempt to measure the eyes of drivers while they were tasked towards inattention out on the open roads (i.e. overt distraction). Instead we took semi-naturalistic observations of both drivers and passengers, and included an experimental tasking period for passengers to act with their eyes as if they were driving for baseline directional verification. Without the availability or safety implication issues of using actual AVs to investigate concerns for lapses in supervisory driver attention, drivers and passengers in conventional vehicles are appealing comparison cases of sometimes near but definitively *“out-of-the-loop”* visual control (i.e., covert inattention) that can evidently produce significantly different eye movement behavior results. Regarding ECC as a distance-based off-center eye measure, the eyes of drivers exhibited greater focus around a central area while the eyes of passengers were more liberal in exploration.

For Phase 2, better signal detection discrimination performance (i.e., increased rates of true positives with decreased rates of false positives), as previously suggested as critical for effective human automation interaction (Lee & See, 2004; Parasuraman & Wickens, 2008), was obtained via accounting for situated driving aspects for a hypothetical DMS. By such an approach, lower amounts of driving-eye data (i.e., smaller windows down to sample level) were observably more susceptible (decreased discriminability between in- vs. out-of-the-loop) to particular driving events/scenarios. This can be seen in our data to occur for both exceedingly high/low driving task demands. On the one hand, increases in driver ECC were seen during an *“urban”* like driving scenario involving low-speed high-degree turns amidst high amounts of potential traffic conflicts. Here, looking around more (at greater off-center distances) might be considered a beneficial adaptive consequence of bandwidth-driven sampling to obtain/maintain a situation awareness across aspects that are (potentially) rapidly changing/fleeting (i.e., information decay). On the other hand, some increases in driver ECC towards levels of out-of-the-loop passengers were seen during a more *“rural”* like driving segment involving high-speed flat steering angles and low/no amount of other vehicles. Here, looking around more (at greater off-center distances) might also be considered a beneficial (rather than mal-adaptive) consequence. Increased road/infrastructure affordances can be considered to effectively protect/contribute increased driving control/predictability (i.e., this portion of our route involved an elongated relatively straight dedicated/segregated expressway off-ramp) and thus the driver could prioritize his/her visual activity/energy to seek additional relevant off-center information (e.g., from signage) or appropriately schedule involvement in a secondary task from left-over/untapped resources (e.g., risk homeostasis). In either case (i.e., our first and second situated insets in Figures 3.3.6-9), such increases in driver ECC could be considered safe/innocuous *ipso facto* as we completed all 32 trips without any perceptible increases in risk of corrective actions, near-crashes, and/or crashes (i.e., definitional components of distraction according to subject matter experts such as described in Hedlund et al., 2006).

If alerts are defined in too absolute rather than relative terms, they run the risk of being overly triggered (i.e., out of context). Too many triggers without actual (or even perceived) necessity for such alerting contributes to false alarms in a *‘cry wolf effect’* and may diminish the effectiveness of a DMS through lowered end-user trust and acceptance. In developing evaluation protocols of emerging DMS technologies, NHTSA (2013b) concluded that *‘perhaps the most important outcome*

of this analysis is an understanding that distraction and its detection cannot be considered independent of the driving environment' while recognizing dependencies between false alarms and acceptance, 'false alarms might either disrupt drivers' attention to the road or undermine their acceptance of the mitigation system'.

4.1 Limitations

Our results regarding eccentricity should be taken with several considerations. In our present analyses, the "pre-task" passenger eye data may have included reading the instructions for some participants as only a gross/generalized cut-off of 20s was applied rather than individually derived. Furthermore, if a driver more fully concentrated on a secondary task to the point of becoming a primary task, it might be expected that eccentricity would decrease rather than increase. Without world knowledge, the eccentricity measure is agnostic as to what specifically is being concentrated on, but instead reflects more only the presence/absence of visual concentration/control. Thus our ECC measure is proposed as only an additional tool to complement AOI-based measures that might supply detections of such overt distraction (e.g., head down and/or eyes directed towards the interior of the vehicle, non-driving related display surfaces, etc.). Future studies should examine the attentional impact of our present situated measures in greater fidelity/detail. For example, vehicle count data was presently determined only by out-of-the-box computer vision emulations provided in MATLAB R2017b Automated Driving System Toolbox via their annotation tool '*Ground Truth Labeler*' strictly by pre-existing automatic processes (ACF Vehicle Detector) without any manual adjustment or deep-learning training modifications. In other words, our vehicle count data is not reflective of state-of-the-art object detection, automated driving scene semantic segmentation, and/or direct-time-of-flight detection (e.g., sonar, radar, lidar, etc.) that might be better situational measurement candidates in future studies. Thus, presently such data should be interpreted for relative precision (repeatability) utility rather than absolute accuracy (validity). Additionally, further contextual aspects of the driving scene for example such as the presence of vulnerable road users (VRU) or real life driving pressures (e.g., driving in haste and/or while in a compromised affectual state) for a more holistic picture of adaptive visual behavior and driving task demands. Lastly, it should be noted that the coordinate frame moved as the participant moved his /her head and it remains for us later outside the scope of the present study to further (re)analyze our data in a rectified/resolved 3D world model as needed.

5. Application

A substantial need for effective DMS is suggested by NHTSA's (2008) crash causation findings primarily consisting of inadequate surveillance, distraction, and inattention. Recent AV developments bring into focus the vigilance dilemmas of drivers turned into supervisors of imperfect self-driving vehicles. In the wake of the first widely reported Tesla Autopilot fatality of Joshua Brown (May 17, 2016 in Florida), the U.S. National Transportation Safety Board (NTSB) issued new safety recommendations on September 12, 2017 for manufacturers to '*develop applications to more effectively sense the driver's level of engagement and alert the driver when engagement is lacking while automated vehicle control systems are in use*'. Meanwhile, accidents with human supervision of so-called 'self-driving' vehicles have continued to occur. On January 30, 2018 in California, a modified Hyundai Genesis by Phantom AI was driving in a supervised autonomy mode and crashed into a lead vehicle in spite of on-board test and press personnel. On March 18, 2018 in California, Elaine Herzberg was killed while walking across the road by an Uber Volvo XC90 while it was driving in an autonomous mode and being supervised by an on-board safety driver.

Many real-time eye tracking DMS algorithms have been developed that can determine if eyes are on/off the road and how frequently the eyes fall within a specified ahead road center region or not. But how '*more effective*' (per the recommendations of NTSB in 2017) both previously developed and released on-road systems may become are often complicated by phenomena such as '*looked-but-failed-to-see*' phenomenon (e.g., in Table 2 of Najm et al., 1994; Hills, 1980; Herslund & Jorgensen, 2003). A Motor Trend review by Hong (2018) characterizes this complication with the camera-based attention monitor of GM's automated driving Super Cruise system as: '*If your eyes are looking forward, but you aren't paying attention ..., this can really catch you out*'. The present analyses have shown how an ECC eye measure can detect visual control aberrance while eyes are still looking up through the windshield. In other words, '*in-the-loop*' vs. '*out-of-the-loop*' eyes of drivers and/or driving automation supervisors can be differentiated from off-center lingering even when the center is not necessarily measured as the road center. Furthermore, reductions in potentially perceived false alarms (e.g., where a DMS might trigger an alert against someone who does not feel distracted as they let their eyes wander during a red light) should help advance the state of the art in DMS towards greater levels of future acceptance and effectiveness.

Acknowledgments

The research presented in this paper was conducted within the project HFAuto – Human Factors of Automated Driving (PITN-GA-2013-605817). Parts of the methods and results of this paper have been previously published in a conference proceedings: Cabral, C.D.D., Petrovych, V., & Happee, R. (2017). Looking at drivers and passengers to inform driver state monitoring of in and out of the loop. In N.A. Stanton (Ed.), *Advances in Human Aspects of Transportation*, Springer International Publishing, pp. 695-707. We would like to thank Dr. Magnus Hjalmdahl for inspiration towards simultaneous eye tracking of driver vs. passenger while driving across Sweden between VTI offices while playing with one of the eye tracking glasses used in the present study. Additionally, we are also indebted to Peter van Leeuwen for his insights into the potential value of driver visual eccentricity behavior, specifically regarding aspects of time and distances away from a center point, and to Dr. Trent Victor for discussions regarding (world-view agnostic) calibrations of gaze center that exploit central tendencies for driver visual re-visits to the same area.

Key points

- We have proposed a measure of eye eccentricity (ECC) that uses median focus areas to calculate off-center looking behavior as a continuous distance-based measure, and it is been shown capable of differentiating eye movements that are "*in-the-loop*" of driving vs. those that are not.
- ECC discrimination between driver and passenger eyes was enhanced by increased amounts of data: breadth-wise through longer moving average windows, as well as depth-wise from increased knowledge and consideration of contextual constraints (i.e., vehicle speed, steering angle, presence of other vehicles).

References

AAA (2018). AAA finds partially automated vehicle systems struggle in real-world conditions. *NewsRoom of the American Automobile Association*, Nov, 14, 2018.
<https://newsroom.aaa.com/2018/11/american-misjudge-partially-automated-driving-systems-ability-based-upon-names/>

Banks, V.A., & Stanton, N.A. (2017). Analysis of driver roles: Modelling the changing role of the driver in automated driving systems using EAST. *Theoretical Issues in Ergonomics Science*. Advanced online publication. doi: 10.1080/1463922X.2017.1305465

Banks, V.A., Eriksson, A., O'Donoghue, J., & Stanton, N.A. (2018). Is partially automated driving a bad idea? Observations from an on-road study. *Applied Ergonomics*, 68, pp. 138-145

CA DMV (2018). Testing of autonomous vehicles with a driver: Adopted regulations for testing of Autonomous vehicles by manufacturers. State of California, Department of Motor Vehicles.
<https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/testing>

Cabral, C.D.D., Petrovych, V., & Happee, R. (2017). Looking at drivers and passengers to inform driver state monitoring of in and out of the loop. In N.A. Stanton (Ed.), *Advances in Human Aspects of Transportation*, Springer International Publishing, pp. 695-707.

Coppola, C., & Frank, B.J. (2018). Report: Uber driver was watching 'The Voice' moments before fatal Tempe crash. Arizona Republic, AZ Central, June 22, 2018.
<https://eu.azcentral.com/story/news/local/tempe-breaking/2018/06/21/uber-self-driving-car-crash-tempe-police-elaine-herzberg/724344002/>

Courage, C., Milgram, P., & Smiley, A. (2000). An investigation of attention demand in a simulated driving environment. *Proceedings of the 14th Triennial Congress of the International Ergonomics Society and 44th Annual Meeting of the Human Factors and Ergonomics Society*. San Diego, California, USA. pp. 3.336-3.339.

Crundall, D.E., & Underwood, G. (1998). Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics*, 41(4), pp. 448-458.

Dinges, D.F., Mallis, M., Maislin, G., & Powell, J.W. (1998). Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management. Department of Transportation Highway Safety, Report No. DOT HS 808 762.

Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2011). Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems*, 12(2), pp. 596-614.

Donmez, B., Boyle, L.N., & Lee, J.D. (2007). Safety implications of providing real-time feedback to distracted drivers. *Accident Analysis and Prevention*, 39, pp. 581-590.

Endsley, M.R., & Jones, D.G. (2012). *Designing for situation awareness: An approach to user-centered design. Second Edition*. Boca Raton, FL, USA: CRC Press, Taylor & Francis Group.

Endsley, M.R., & Rodgers, M.D. (1998). Distribution of attention, situation awareness, and workload in a passive air traffic control task: Implications for operational errors and automation. *Air Traffic Control Quarterly*, 6(1), pp. 21-44.

Endsley, M.R., Mogford, R., Allendoerfer, K., Snyder, M.D., & Stein, E.S. (1997). *Effect of free flight conditions on controller performance, workload, and situation awareness: A preliminary investigation of changes in locus of control using existing technology*. (DOT/FAA/CT-TN 97/12). Atlantic City, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Euro NCAP (2018). For the first time, Euro NCAP puts automated driving technology to the test. *Euro NCAP, Press Releases*, October 18, 2018.
<https://www.euroncap.com/en/press-media/press-releases/testingautomation/>

Godthelp, J., Milgram, P., Blaauw, G.J. (1984). The development of a time-related measure to describe driving strategy. *Human Factors*, 26(3), pp. 257-268.

Gordon, D.A. (1966). Experimental isolation of the driver's visual input. *Human Factors*, 8(2), pp. 129-137.

Gray, R., & Regan, D.M. (2005). Perceptual processes used by drivers during overtaking in a driving simulator. *Human Factors*, 47(2), pp. 394-416.

Green, P.A. (2001). Where do drivers look while driving (and for how long)? In *Human Factors in Traffic Safety*. Tucson, Lawyers and Judges Publishing Company, Chap. 4, pp. 77-110.

Greenlee, E.T., DeLucia, P., & Newton, D.C. (2018). Driver vigilance in automated vehicles: Hazard detection failures are a matter of time. *Human Factors*, 60(4), pp. 465-476.

Hancock, P.A. (2013). Driven to distraction and back again. In M.A. Regan, J.D. Lee, & T. Victor (Eds.), *Driver distraction and inattention: Advances in research and countermeasures*. Surrey, UK: Ashgate, pp. 9- 26.

Hedlund, J., Simpson, H., & Mayhew, D. (2006). International conference on distracted driving:

Summary of proceedings and recommendations (2-5 October). Retrieved from <http://www.distracteddriving.ca/english/documents/ENGLISH-DDProceedingsandRecommendations.pdf>

Heikoop, D.D., De Winter, J.C.F., van Arem, B., & Stanton, N.A. (2015). Psychological constructs in driving automation: A consensus model and critical comment on construct proliferation. *Theoretical Issues in Ergonomics Science*, 17(3), pp. 284-303.

Herslund, M.B., & Jorgensen, N.O. (2003). Looked-but-failed-to-see-errors in traffic. *Accident Analysis & Prevention*, 35(6), pp. 885-891.

Hills, H.L. (1980). Vision, visibility, and perception in driving. *Perception*, 9(2), pp. 183-216.

Hong, P. (2018). 2018 Cadillac CT6 Super Cruise: Can the car really drive itself? We take the semi autonomous system for a 1,000-mile drive to test its self-steering adaptive cruise control. Motor Trend Buyer's Guide review, April 24, 2018. <http://www.motortrend.com/cars/cadillac/ct6/2018/2018-cadillac-ct6-super-cruise-review/>

Ji, Q., & Yang, X. (2002). Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real Time Imaging*, 8, pp. 357-377.

Just, M.A., & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), pp. 329-354.

Kircher, K., & Ahlstrom, C. (2009). Issues related to the driver distraction detection algorithm AttenD. *Proceedings of the 1st International Conference on Driver Distraction and Inattention*. Gothenburg, Sweden. pp. 1-15.

Kircher, K., & Ahlstrom, C. (2017). Minimum required attention: A human-centered approach to driver inattention. *Human Factors*, 59(3), pp. 471-484.

Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., & Ramsey, D.J. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. NHTSA, Washington, DC., Technical Report no. DOT-HS-510-594.

Land, M.F. & Lee, D.N. (1994). Where we look when we steer. *Nature*, 369, pp. 742-744.

Laya, O. (1992). Eye movements in actual and simulated curve negotiation tasks. *IATSS Research*, 16(1), pp. 15-26.

Lee, J.D. (2008). Fifty years of driving safety research. *Human Factors*, 50(3), pp. 521-528.

Lee, J.D., & See, K.A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), pp.50-90.

Liang, Y., Lee, J.D., & Yekhshatyan, L. (2012). How dangerous is looking away from the road? Algorithms predict crash risk from glance patterns in naturalistic driving. *Human Factors*, 54(6), pp. 1104-1116.

Louw, T., & Merat, N. (2017). Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transportation Research Part C: Emerging Technologies*, 76, pp. 35-50.

Louw, T., Madigan, R., Carsten, O., & Merat, N. (2016). Were they in the loop during automated driving? Links between visual attention and crash potential. *Injury Prevention*, 23(4), pp. 1-6.

Mackworth, N. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1:1, p. 6-21.

Madhaven, P., Wiegmann, D.A., Lacson, F.C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors*, 48(2), pp. 241-256.

Mays, K. (2018). Which cars have self-driving features for 2018? Cars.com, April 20, 2018. <https://www.cars.com/articles/which-cars-have-self-driving-features-for-2018-1420699785509/>

Merat, N., Jamson, A.H., Lai, C.H., Daly, M., & Carsten, O.M.J. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behavior*, 27, pp. 274-282.

Merat, N., Seppelt, B., Louw, T., Engstrom, J., Lee, J.D., Johansson, E., Creen, C.A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., & Keinath, A. (2018). The 'out-of-the-loop' concept in automated driving: proposed definition, measures and implications. *Cognition, Technology & Work*. <https://doi.org/10.1007/s10111-018-0525-8>

Metzger, U., & Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control vs. passive monitoring. *Human Factors*, 43(4), pp. 519-528.

Mourant, R.R., & Ge, Z. (1997). Measuring attentional demand in a virtual environments driving simulator. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*. Albuquerque, New Mexico, USA. pp. 1268-1272.

Mourant, R.R., & Rockwell, T.H. (1970). Mapping eye-movement patterns to the visual scene in driving: An exploratory study. *Human Factors*, 12(1), pp. 81-87.

Mourant, R.R., Rockwell, T.H., & Rackoff, N.J. (1969). Drivers' eye movements and visual workload.

Washington, DC: Highway Research Record, No. 292

Najm, W.G., Koziol Jr., J.S., Tijerina, L., Pierowicz, J.A., & Hendricks, D.L. (1994). Comparative assessment of crash causal factors and IVHS countermeasures. *Proceedings of the IVHS America Annual Meeting*. Atlanta, Georgia, USA.

NHTSA (2008). National motor vehicle crash causation survey, Report to Congress. (Report No. DOT HS 811 059). Washington, DC: National Highway Traffic Safety Administration (NHTSA).

NHTSA (2013a). Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices. United States National Highway Traffic Safety Administration (NHTSA). Docket No. NHTSA-2010-0053. *Federal Register*, 78(81), April 26, 2013, Notices.

NHTSA (2013b). Distraction detection and mitigation through driver feedback. United States National Highway Traffic Safety Administration (NHTSA), Report No. DOT HS 811 547A

NHTSA (2017). Automated driving systems 2.0: A vision for safety. Introductory Message from U.S. Department of Transportation Secretary, Elaine L. Chao, Guidance and Best Practices from U.S. Department of Transportation.

Ohn-Bar, E., & Trivedi, M.M. (2016). Looking at humans in the age of self-driving and highly automated vehicles. *IEEE Transactions on Intelligent Vehicles*, 1(1), pp. 90-104.

Olson, P.L. (1964). The driver's reference point as a function of vehicle type, direction and radius of turn. *Human Factors*, 6(31), pp. 319-325.

Olson, P.L., Battle, D.S., Aoki, T. (1989). Driver eye fixations under different operating conditions. University of Michigan Transportation Research Institute, Report No. UMTRI-89-3, Industry Affiliation Program for Human Factors in Transportation Safety.

Pampel, S.M., Large, D.R., Burnett, G., Matthias, R., Thompson, S., Skrypchuk, L. (2018). Getting the driver back in the loop: The quality of manual vehicle control following long and short non-critical transfer-of-control requests: *TI:NS. Theoretical Issues in Ergonomics Science*. (in press).

Parasuraman, R. (1987). Human-computer monitoring. *Human Factors*, 29(6), pp. 695-706.

Parasuraman, R., & Wickens, C.D. (2008). Humans: Still vital after all these years of automation. *Human Factors*, 50(3), pp. 511-520.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, 38 (4), pp. 665-679.

Rackoff, N.J., & Rockwell, T.H. (1975). Driver search and scan patterns in night driving. *Transportation Research Board Special Report*, 156, pp. 53-63.

Recarte, M.A., & Nunes, L. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied*, 6(1), pp. 31-43.

Reimer, B. (2009). Impact of cognitive task complexity on drivers' visual tunneling. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2138, Transportation Research Board of the National Academies, Washington, DC. Pp 13-19.

Reyes, M.L., & Lee, J.D. (2008). Effects of cognitive load presence and duration on driver eye movements and event detection performance. *Transportation Research Part F: Traffic Psychology and Behavior*, 11, pp. 391-402

Rockwell, T.H. (1988). Spare visual capacity in driving – revisited: New empirical results of an old idea. In A.G. Gale, M.H. Freeman, C.M. Haslegrave, P. Smith, & S.P. Taylor (Eds.). *Vision in Vehicles II*. Amsterdam, Elsevier Science, pp. 317-324.

Romoser, M.R.E. (2013). The long-term effects of active training strategies on improving older drivers' scanning in intersections: A two-year follow-up to Rosmoser and Fisher (2009). *Human Factors*, 55(2), pp. 278-284.

SAE (2018). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. SAE International, Surface Vehicle Recommended Practice, J3016_201806. Superseding J3016 SEP2016.
https://www.sae.org/standards/content/j3016_201806/

Salvucci, D.D., & Gray, R. (2004). A two-point visual control model of steering. *Perception*, 33, pp. 1233-1248.

Schmidt, E.A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., Kincses, W.E. (2009). Drivers' misjudgment of vigilance state during prolonged monotonous driving. *Accident Analysis and Prevention*, 41, pp. 1087-1093.

Senders, J.W., Kristofferson, A.B., Levison, W.H., Dietrich, C.W., & Ward, J.L. (1967). Attentional demand of automobile driving (Technical Report No. 1482). Cambridge, MA: Bold, Beranek, and Newman, Inc.

Serafin, C. (1994). Driver eye fixations on rural roads: Insight into safe driving behavior. University of Michigan Transportation Research Institute, Report No. UMTRI-94-21, Federal Highway Administration.

Sheridan, T.B. (2004). Driver distraction from a control theory perspective. *Human Factors*, 46(4), pp. 587-599.

Shinar, D., McDowell, E.D., & Rockwell, T.H. (1977). Eye movements in curve negotiation. *Human Factors*, 19(1), pp. 63-71.

Shojaeizadeh, M., Djamasbi, S., & Trapp, A.C. (2016). Density of gaze points within a fixation and information processing behavior. In *International Conference on Universal Access in Human-Computer Interaction*. Springer International Publishing. pp. 465-471.

Sivak, M. (1996). The information that drivers use: It is indeed 90% visual? *Perception*, 25(9), pp. 1081-1089.

Smith, P., Shah, M., & da Vitoria Lobo, N. (2000). Monitoring head/eye motion for driver alertness with one camera. *The 15th International Conference on Pattern Recognition*, 4, pp. 636-642.

Takeda, Y., Sato, T., Kimura, K., Komine, H., Akamatsu, M., & Sato, J. (2016). Electrophysiological evaluation of attention in drivers and passengers: Toward an understanding of drivers' attentional state in autonomous vehicles. *Transportation Research - Part F: Traffic Psychology and Behavior*, 42, pp. 140-150.

Tijerina, L., Barickman, F.S., Mazzae, E.N. (2004). Driver eye glance behavior during car following. National Highway Traffic Safety Administration, Technical Report No. DOT HS 809 723.

Vicente, F., Huang, Z., Xiong, X., de la Torre, F., Zhang, W., & Levi, D. (2015). Driver gaze tracking and eyes off the road detection system. *IEEE Transactions on Intelligent Transportation Systems*, 16(4), pp. 2014-2027.

Victor, T., Harbluk, J.L., Engstrom, J.A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behavior*, 8, pp. 167-190.

Weller, G. & Schlag, B. (2010). A robust method to detect driver distraction. *Proceedings of the 2nd International HUMANIST Virtual Centre of Excellence Conference*. Berlin, Germany.

Wierwille, W. (1993). Visual and manual demands of in-car controls and displays. In B. Peacock & W. Karwowski (Eds.), *Automotive ergonomics*. New York: Taylor & Francis, pp. 299-320.

Yoshida, J. (2018). Driver Monitoring: 'Shut up, Mom, I see it!'. *Electronic Engineering Times*, April 30, 2018. https://www.eetimes.com/document.asp?doc_id=1333236

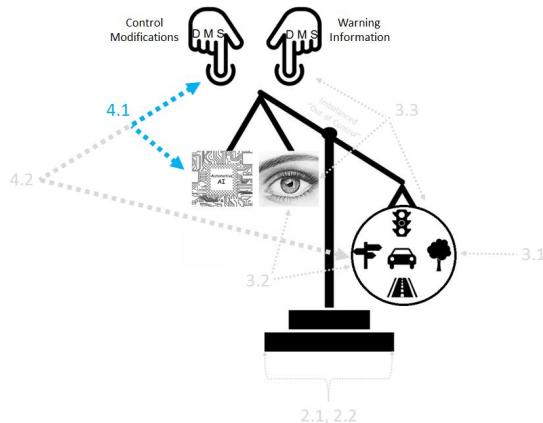
Zhang, H., Smith, M.R.H., & Witt, G.J. (2006). Identification of real-time diagnostic measures of visual distraction with an automatic eye-tracking system. *Human Factors*, 48, pp. 805-821.

Zhang, H., Smith, M.R.H., & Dufour, R. (2008). A final report of safety vehicles using adaptive interface technology (Phase II: Task 7C): Visual Distraction.

4

PART 4: Adaptive Driving Automation

Chap. 4.1) Adaptive automation: Automatically (dis)engaging automation during visually distracted driving



In regards to the overall thesis big picture, this driving simulator experiment serves to provide initial validations of integrating a real-time eye-based driver monitoring system with driving automation system functionality. As a result, lateral performance was observably improved in a visual distraction induced backup driving automation system compared to conventional/manual controlled driving with the same visual distractions. Eye tracking was used here on an applications basis but not investigated as a primary research factor of interest. Instead, the adaptive directionality of automatic consequential vehicular control transfer was varied to either end up with the human or the automation upon detection of visual distraction. Participants performed better with (less lateral error) and better appreciated (lower workload and higher acceptance ratings) the backup concept. Chap. 4.2 extended the successful backup driving automation of Chap. 4.1 with enhanced simulation visual/behavioral fidelity as well as an investigation of further design aspects of the DMS integration to address potential human interaction drawbacks of over-alerting and over-reliance.

Adapted from:

Cabral, C.D.D.*; Janssen, N.M.*; & de Winter, J.C.F. (2018). Adaptive automation: Automatically (dis)engaging automation during visually distracted driving. *PeerJ Computer Science*, 4:e166, <https://doi.org/10.7717/peerj-cs.166>

*Joint first authors

Abstract

Background: Automated driving is often proposed as a solution to human errors. However, fully automated driving has not yet reached the point where it can be implemented in real traffic. This study focused on adaptively allocating steering control either to the driver or to an automated pilot based on momentary driver distraction measured from an eye tracker.

Methods: Participants (N = 31) steered a simulated vehicle with a fixed speed, and at specific moments were required to perform a visual secondary task (i.e., changing a CD). Three conditions were tested: (1) Manual driving (Manual), in which participants steered themselves. (2) An automated backup (Backup) condition, consisting of manual steering except during periods of visual distraction, where the driver was backed up by automated steering. (3) A forced manual drive (Forced) condition, consisting of automated steering except during periods of visual distraction, where the driver was forced into manual steering. In all three conditions, the speed of the vehicle was automatically kept at 70 km/h throughout the drive. **Results:** The Backup condition showed a decrease in mean and maximum absolute lateral error compared to the Manual condition. The Backup condition also showed the lowest self-reported workload ratings and yielded a higher acceptance rating than the Forced condition. The Forced condition showed a higher maximum absolute lateral error than the Backup condition. **Discussion:** In conclusion, the Backup condition was well accepted, and significantly improved performance when compared to the Manual and Forced conditions. Future research could use a higher level of simulator fidelity and a higher-quality eye-tracker.

1. Introduction

1.1. Automated driving

Over the last couple of decades, researchers have been studying the viability of automated driving for commercial use. However, automation research has not yet reached the point where fully autonomous driving can be implemented with the promise of a perfect system. Current designs of automated driving systems often focus on applying partial, conditional, or high automation (SAE International, 2016), where the human tasks are that of a supervisor. This supervisory role has brought about other human factor issues, including loss of vigilance, varying workload, fatigue, and loss of situation awareness (Casner, Hutchins & Norman, 2016; De Winter et al., 2014; Matthews, 2016; Parasuraman & Riley, 1997).

1.2. The potential of adaptive automation

Adaptive automation has been proposed as a solution to maximize human-machine cooperation (e.g., De Visser & Parasuraman, 2011; Hancock, 2007; Inagaki, 2003; Kaber & Endsley, 2004; Parasuraman, 2000). In adaptive automation, control functions change to a lower or higher level of automation depending on predetermined criteria, such as momentary workload or situation awareness of the human operator. For example, if during the automated execution of a task the human is measured to be inattentive, the algorithm could switch the automation to a lower level or even turn over control entirely to the human operator to engage the human. Alternatively, if high human workload is detected during a manually-executed task, some or all of the control might automatically be switched to the automation.

1.3. Types of adaptive automation

Algorithms that define how and when automation is invoked and terminated differ greatly. Sheridan & Parasuraman (2005) (see also Parasuraman et al., 1992; Inagaki, 2003) describe five types of methods for implementing adaptive automation: (1) critical-event logic, (2) operator performance measurements, (3) modeling, (4) operator physiological measurements, and (5) hybrid methods, combining multiple of these methods.

Physiological measurements offer the advantage that they can be obtained continuously regardless of whether the automation is active or inactive (Parasuraman et al., 1992; Scerbo et al., 2001). There are different physiological measures that provide information on the human operator state, including heart rate, skin conductance, and eye movements. For this research study, the focus lies on eye movements because they can be measured non-obtrusively and provide specific information regarding where the driver attends to, as opposed to other physiological indexes which provide a more general index of attentional/arousal. How drivers distribute their visual attention is relevant in driving safety research, as the information relevant to driving is likely to be predominantly visual (Sivak, 1996).

1.4. Backup automation

Fundamentally, there are two approaches towards adaptive automation using eye movements. The first approach is backup or background automation, which “allows the driver to drive the vehicle, but watches over them in case of trouble” (Kyriakidis et al., in press). For example, it is possible to let the driver control the car manually, and invoke automation if the driver is distracted. This

approach may be beneficial for safety, as off-road glances are associated with decrements in performance and safety. For example, a naturalistic driving study found significant associations between eyes-off-road time and standard deviation of lateral position (Peng, Boyle & Hallmark, 2013).

Backup automation is similar to real-time distraction-mitigation feedback which alerts drivers based on their off-road eye glances (Donmez, Boyle & Lee, 2007). However, alerts alone may not always be effective, as drivers may decide to ignore warning systems (e.g., Parasuraman & Riley, 1997).

1.5. *Forced manual driving*

The second and opposite approach ('foreground automation'; Kyriakidis et al., *in press*) would be to let the car drive automatically, and force the driver to take over if he or she is distracted (i.e., negligent in their responsibility for monitoring the dynamic driving task).

The notion of forced manual driving might seem odd due to its apparent unsafe nature. However, it is not odd in the sense that it roughly corresponds to a path being followed by the automotive industry. As a result of an investigation into the first fatal crash with Tesla's Autopilot and a truck in May 2016, the National Transportation Safety Board (NTSB) has issued recommendations to "develop applications to more effectively sense the driver's level of engagement" and to "incorporate system safeguards that limit the use of automated vehicle control systems to those conditions for which they were designed" (NTSB, 2017). Other than a warning based on hands-on-wheel sensing, one such safeguard could be to automatically activate a functional transition from the automated mode towards manual driving, see the case of Cadillac Super Cruise, which uses head tracking software that "helps make sure your eyes are on the road, and alerts you when you need to pay more attention or take back control" (Cadillac, 2018). In current level 2 automated driving, the car performs lateral and longitudinal control, and the system penalizes the inattentive supervisor with a transition of control back to the driver. In an overview of 2017 models from vehicle manufacturers with level 2 driving automation systems, transitions of control back to the driver were found to be a commonly employed strategy for reacting to insufficient supervisory driver attention (C Cabrall, A Eriksson, F Dreger, R Happee & JCF De Winter, 2018, unpublished data). Accordingly, the forced manual driving may be a useful strategy to prevent overreliance on automation.

1.6. *The present study*

In summary, transitions in adaptive automation could occur in two directions. While driving manually, detection of visual distraction could trigger a transition from manual driving control to automated control (Backup automation). In the other direction, visual distraction could trigger a transition from automated to manual driving (Forced manual driving). At present, it is unknown whether background automation or foreground automation with forced manual driving is preferred in terms of safety and driver acceptance.

The present experiment was performed with three different conditions (1) Manual driving (Manual), (2) An automated backup (Backup) condition, consisting of manual driving except during periods of visual distraction, where the driver was backed up by an automated pilot that was automatically initiated, and (3) A forced manual drive (Forced) condition, consisting of automated

driving except during periods of visual distraction, where the driver was forced back into the manual control loop.

An expected result was that the automated backup condition would yield better lane-keeping performance during visual distraction because the automation is programmed to keep lane center better than what humans are capable of. Additionally, it was of interest to see whether people accepted this condition, in which control was taken away from them. For the forced manual drive condition, it was expected that lateral driving performance would deteriorate as compared to the manual drive condition during such moments because visual attention is a prerequisite for being able to keep the car in the lane (Senders et al., 1967).

2. Methods

2.1. Ethics statement and Participants

This research was approved by the Human Research Ethics Committee (HREC) of the Delft University of Technology (TU Delft). All participants provided written informed consent. Thirty-one people participated, of which 25 were male and six female. The mean age was 26.4 years ($SD = 4.5$ years). Participation criteria were having a driver's license, and not having to wear glasses to see properly. Participants were offered €5 compensation for their time (approx. 30 min).

2.2. Equipment

A SmartEye DR120 remote eye tracker was used to record the participant's gaze direction while seated and viewing a desktop monitor (Figure 4.1.1). Data were collected at a frequency of 60 Hz. The experiment took place in a room with standard office lighting and lowered window blinds. A 24-inch monitor was used to display the simulated environment. The distance between the monitor and the participant differed between participants but was limited by the DR120 eye tracker, which was able to measure in the range 50–80 cm from the cameras. A Logitech G27 steering wheel was used to control the simulated vehicle. PreScan software (TASS International, Helmond, The Netherlands) was used to create the simulation environment. MATLAB/Simulink was used along with PreScan to control the simulated vehicle and to log data. A stack of CDs and a small boom box to the right of the monitor and steering wheel were used to present a secondary task that evokes visual distraction similar to that which might commonly occur while driving (e.g., using a route navigation device, tuning the radio, texting).



Figure 4.1.1. The experimental setup.

2.3. Simulated environment

The environment consisted of a two-lane road with a lane width of 5 m. The road had five straight segments and four 10° bends (Figure 4.1.2). The participant was shown the dashboard of a vehicle (BMW X5) as well as the road in front of them (Figure 4.1.3). A bar on the dashboard indicated the state of the automation. A green bar indicated that the automation was on, a yellow bar indicated that the automation was still on but that the participant was about to regain lateral control, and a red bar indicated that the automation was off (i.e., manual lateral control). The automation was designed in such a way that when it was switched on, it would quickly drive the car towards the center of the right lane and keep it there.

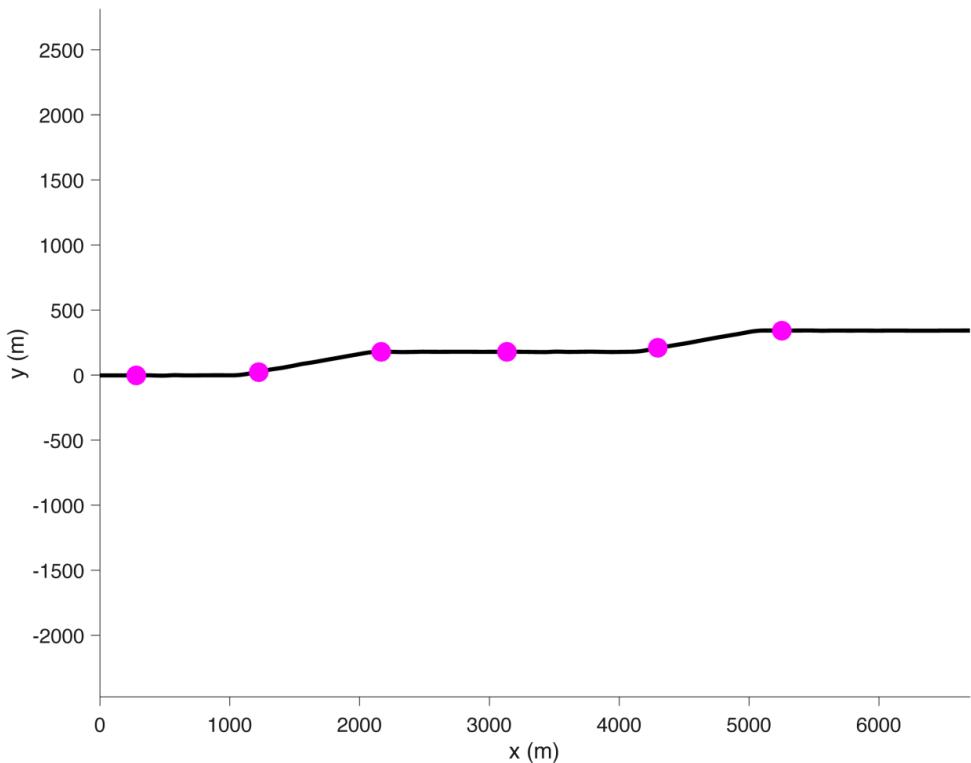


Figure 4.1.2. Top-down perspective of the road. The markers indicate the six moments when a 1-s beep was presented, signaling that the participant could start the secondary task.

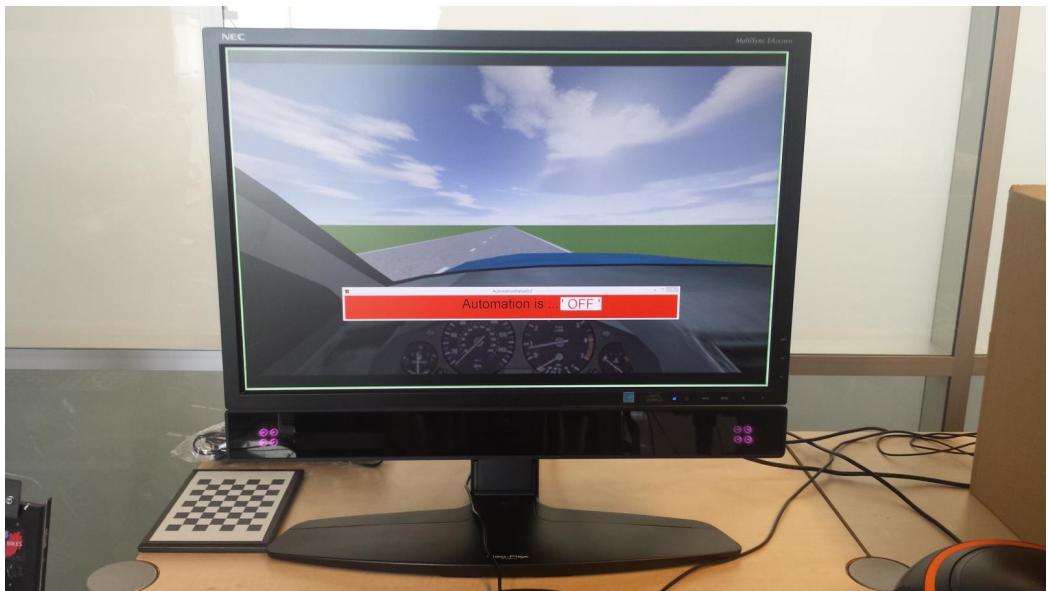


Figure 4.1.3. Photo from the participant's perspective. The eye-tracker cameras are connected to the bottom of the monitor.

2.4. Experimental conditions

A within-subject design was used, and the order of the conditions was counterbalanced across the participants. The counterbalancing was done by presenting the six possible orders of the Manual (1), Backup (2), and Forced (3) conditions, in the following manner to the first six participants: 1-2-3, 3-2-1, 2-3-1, 2-1-3, 1-3-2, 3-1-2. These orders were repeated for Participants 7–12, 13–18, 19–24, and 25–30, and Participant 31 was presented with the 1-2-3 order. During the entire experiment, the vehicle speed was constant at 70 km/h, and thus no longitudinal control actions were required. This speed was chosen to simulate driving on rural roads. No infrastructure (buildings, signage, vegetation) nor any other traffic were simulated. Three experimental conditions were used:

- (1) Manual condition (Manual). In the Manual condition, the participant performed the steering without help from an automated system.
- (2) Automated backup condition (Backup). In this condition, the automated system assumed lateral control when visual distraction was measured. Otherwise, the participant performed manual steering. Visual distraction was defined by the consecutive eyes-off-monitor time being greater than 1.5 s. The secondary task was placed to the right of the steering wheel, and when the participant turned the head to look at it, it sometimes became difficult for the eye tracker to record the eyes. When the eye tracker was not able to record the eyes, it reported this as null values, and the algorithm treated these as off-monitor measurements. Automation termination was also performed based on eye measurements: the participant would regain lateral control if (s)he focused on the monitor for 4.5 s. The yellow status bar switched on 1.4 s before the transition to manual took place. The 1.5 s and 4.5 s thresholds were based on pilot studies (see <https://data.4tu.nl/repository/uuid:49d87edc-07a6-4f07-a5e6-0b699705881b>). The 1.5 s threshold for Backup automation is in approximate agreement with the literature, which suggests that off-road glances of 2.0 s and longer are risky (Klauer et al., 2006; Ryu, Sihn & Yu, 2013). Recently, Liang, Lee & Horrey (2014) concluded that “frequent off-road glances longer than 1.7 s present a high-risk glance pattern in the seconds preceding a safety-critical event and that the 2.0 second-threshold that is frequently cited in defining dangerously long off-road glances might be a liberal estimation”.
- (3) Forced manual drive condition (Forced). The Forced condition can be described as being opposite to the Backup condition in the sense of control transition directionality. The automation had lateral control of the car while the participant was assessed as being visually attentive, and initiated a control transition to manual driving if visual distraction was measured. If the gaze was directed away from the monitor for 1.5 consecutive seconds, the automation switched off, and the participant would be forced to drive manually. The status bar switched from green to yellow 0.75 before the transition to manual would take place. The algorithm would wait until 4.5 on-monitor seconds were measured and then the automation would switch on.

In the Manual and Backup conditions, the first 3.5 s of each trial were driven with automation enabled, and between 3.5 and 5 s, the status bar was yellow. This ensured that the participant started smoothly with zero lateral error.

During automated driving, the steering wheel (i.e., the physical angle of the Logitech steering wheel) was decoupled from the simulated steering angle, and so not necessarily centered. When regaining manual control, the virtual steering angle would make a discrete jump from the previous steering angle determined by the automation towards the steering angle at which the physical steering wheel angled at that moment.

It is noted that the above-mentioned descriptions of the Backup and Forced conditions are simplifications of the actual algorithms (see <https://doi.org/10.4121/uuid:49d87edc-07a6-4f07-a5e6-0b699705881b> for source code). One detail is that, to prevent effects of eye blinks and rapid glances between the secondary task and the monitor, the algorithms featured a filter regarding the transition back to the nominal state (i.e., manual driving in the Backup condition and automated driving in the Forced condition). This means that the driver did not have to look to the monitor for 4.5 s consecutively to induce a transition. Specifically, the algorithm of the Backup condition was programmed in such a way that if the eye tracker measured 1.5 consecutive off-monitor seconds, the on-monitor counter would reset to zero. In other words, if a cumulative total of 4.5 on-monitor seconds were measured (i.e., without 1.5 consecutive off-monitor seconds in between), the participant automatically regained lateral control. For the Forced condition, on the other hand, the on-monitor counter would reset after 0.33 consecutive off-monitor seconds. There was no specific purpose for these differences between the Backup and Forced conditions, but these differences were the consequence of adjustments during pilot testing.

2.5. Secondary task

The participant was given a secondary task intended to cause a visual distraction. In this secondary task, the participant was required to perform a sequence of physical actions involving the stack of CDs and a CD-player (see Horberry et al., 2006), who reported that this type of task degrades driving performance).

The sequence of steps consisted of keeping the left hand on the steering wheel and using the right hand to (1) press stop on the CD-player, (2) open the CD-player, take out the CD, and put it on top of the stack of CDs, (3) take out the bottom CD from the stack, put it in the CD-player, and close the lid, (4) press play on the CD-player, (5) put the stack of CDs back in their original position, and (6) place the right hand back on the steering wheel. The sequence of steps was designed to encourage visual distraction and thus trigger an automatic transition of control. Note that the volume of the CD-player was set to zero.

The participant was told to keep the left hand on the steering wheel at all times. Furthermore, the participant was instructed to look at the secondary task (CDs, CD-player) when performing the secondary task. In other words, the participant was not supposed to look towards the monitor and simultaneously perform the secondary task based on peripheral vision or touch. This requirement was included to ensure that the participant was visually distracted from the driving due to performing the secondary task.

At six moments during the drive (after 15 s, 65 s, 115 s, 165 s, 225 s, and 275 s), the participant was alerted that he/she was required to perform the secondary task by a long (1 s) beep. To encourage secondary task engagement, the participant was scored by the experimenter on a scale from 0 to 10. The participant could get up to 6 points for performing the task steps correctly and up to 4 points depending on how quickly the task was completed. The scoring was done by the

experimenter by looking at the participant and an on-screen timer that was visible on the experimenter's computer. The precise scoring criteria are provided in <https://doi.org/10.4121/uuid:49d87edc-07a6-4f07-a5e6-0b699705881b>. If the task was not completed within 25 s, the participant would only get points for the steps finished at that time. The total score was the average of the six secondary tasks per driving trial. At the end of each driving trial and before they started with the questionnaires, the experimenter orally told the participant what the secondary task score was, rounded to 1 decimal point.

Additionally, for the Manual and Backup conditions, a short (0.25 s) beep was produced 25 s after the long beep, to mark the end of the secondary task period. In the Forced condition, the short beep was produced when the automation had made a transition from manual to automated driving or 25 s after the long beep (whichever came first). For the Forced condition, the short beep was presented right after the manual-to-automation transition to signal to the participant that the secondary task was over.

The instructions form mentioned that "a long beep will indicate the start of the task, a short beep will indicate that you can stop the task if you are not already finished." Furthermore, the form stated that lane keeping was the primary task, "Your primary task is to focus on staying in the center of the right lane as accurately as you can. This should always be the most important task. Safety first!". The form also clarified that changing the CD was the secondary task, and that the participant should attempt to score as high as possible while still driving safely.

2.6. Procedure

After reading and signing the consent form, which mentioned the goal of the experiment and the workings of the three conditions, each participant was asked to fill out a personal information questionnaire. They were also required to read the instructions form (see <https://doi.org/10.4121/uuid:49d87edc-07a6-4f07-a5e6-0b699705881b>).

Next, the participant was asked to sit in front of the eye tracker and focus on four fixed points on the monitor to perform a gaze calibration. If the calibration could not be completed, the participant was asked to sit differently so that the cameras could record their eyes better before performing another calibration.

For each of the three conditions, the participant was asked to drive the simulated vehicle in the environment described above, using the steering wheel for lateral control. Additionally, for each of the three conditions, at fixed intervals during driving, the participant was required to perform the CD-player secondary task.

After each driving trial, the participant was asked to complete a NASA Task Load Index (TLX) questionnaire (Hart & Staveland, 1988). Following the Backup and Forced conditions, the participant was required to fill out an acceptance scale of in-vehicle technology (Van der Laan, Heino & De Waard, 1997). The participants were not required to complete this questionnaire for the Manual condition, because the scale asks to rate a specific vehicle technology. At the end of the experiment, the participant was asked to complete a questionnaire where they could state which

session they preferred as well as give general comments (for all the questionnaires used in this study, see <https://doi.org/10.4121/uuid:49d87edc-07a6-4f07-a5e6-0b699705881b>).

The participant performed a 185 s training run before each of the driving trials to become familiar with each condition. These training runs were driven on the same track as the actual experimental runs, and included three secondary task periods. After the training run, the participant drove the full track, which took 350 s for each driving trial and included six secondary task periods.

2.7. Dependent variables

The following measures and measurements were assessed across the 10.0 s and 349.5 s of elapsed time per driving trial of a particular condition. The first 10 s were discarded because this period was regarded as settling time for participants.

Lateral performance:

- (1) Mean Absolute Lateral Error (meanALE) (m). This was the mean of the absolute difference in lateral position between the vehicle's position and the lane center. The meanALE is an index of *overall* lane keeping performance and includes both periods where the lateral driving automation is active (and so the lateral error is 0) and periods of manual driving.
- (2) Mean Absolute Lateral Error during Manual Driving (meanMALE) (m). This was the mean of the absolute difference in lateral position between the vehicle's position and the lane center, only for moments when the participant was driving manually.
- (3) Maximum Absolute Lateral Error (maxALE) (m). maxALE is the *maximum* of the absolute difference in lateral position between the vehicle's position and the lane center in meters, and can be regarded as an index of safety.

Furthermore, the following measures were extracted from the self-reports, for each of the three driving conditions.

Secondary task performance:

- (4) The secondary task score (0–10) was computed as the mean of the full set of six secondary tasks of a driving trial.

Workload:

- (5) NASA-TLX (%), ranging from 0% to 100% with steps of 5%. This questionnaire was used to assess subjective workload on six different categories: (1) Mental demand, (2) Physical demand, (3) Temporal demand, (4) Performance, (5) Effort, and (6) Frustration (Hart & Staveland, 1988). The items were answered on a 21-point scale ranging from 'very low' ('perfect' for the performance item) to 'very high' ('failure' for the performance item). A composite score was obtained by taking the mean of the six different sub-category scores (Byers, Bittner Jr & Hill, 1989).

System acceptance:

- (6) Acceptance scale, ranging between +2 and -2, with steps of 1. The acceptance scale was used to assess the drivers' opinion on the Usefulness and the Satisfaction of the systems they tested. This questionnaire consisted of nine sub-scale items, presented in order as (1) useful-useless, (2) pleasant-unpleasant, (3) bad-good, (4) nice-annoying, (5) effective-superfluous, (6) irritating-likeable, (7) assisting-worthless, (8) undesirable-desirable, (9) raising alertness-sleep inducing.
- (7) Preference. The participant was also asked which condition they preferred the most in a final questionnaire after they had performed all of the conditions. The question they were asked was "Which session did you prefer?". The possible answers were "session 1", "session 2" "session 3", and "no difference".

2.8. Statistical analyses

Non-parametric tests were used because some of the performance measures were non-normally distributed among participants. For example, maxALE represents the maximal deviation during the entire drive and so is sensitive to a single road excursion. Differences between pairs of conditions were compared using the Wilcoxon signed rank test. Corresponding effect sizes were calculated as $Z/N^{0.5}$. A significance level of .005 was used (Benjamin et al., 2017).

3. Results

3.1. Automation functionality

First, we assessed whether the Backup and Forced conditions worked as intended. Figure 4.1.4 shows the proportion of participants with automation on at any time for the Backup and Forced conditions. It can be seen that about 90% of the participants in the Backup condition drove automatically about 10 s after the task initiation beep was presented. The 10 s comprises the minimum 1.5 s required to initiate a transition, plus individual differences in eye-response time (or the fact that participants may have used frequent scanning back and forth scanning rather than a direct re-allocation of gaze in a binary manner). Similarly, about 90% of the participants were issued manual driving control status in the Forced condition about 10 s after the beep. Figure 4.1.4 also shows that some of the participants experienced control transitions outside of the secondary task periods. This could be due to eye tracker imperfections, as faulty measurements could result in 1.5 s off monitor glancing. Summarizing, the results in Figure 4.1.4 show that the Backup and Forced conditions worked in opposite ways, as intended.

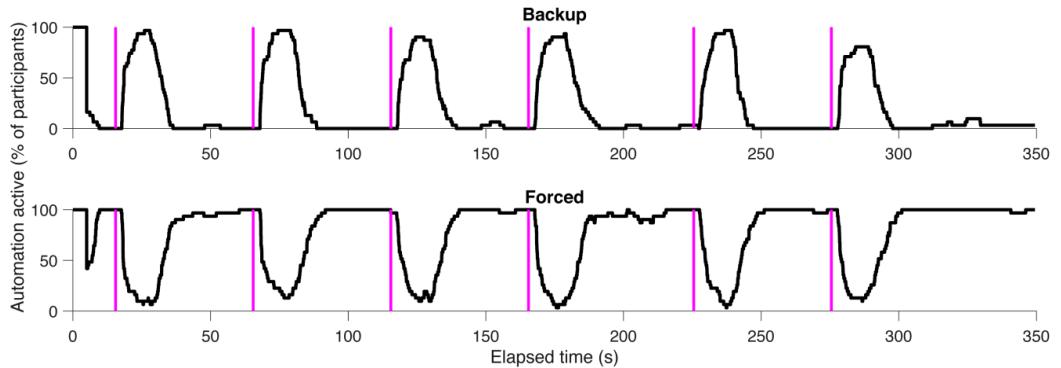


Figure 4.1.4: The proportion of participants with automation turned on as a function of elapsed time. The magenta vertical lines represent the secondary task initiation beeps. The overall percentage of automated driving time was 0%, 22%, and 76% for the Manual, Backup, and Forced conditions, respectively.

3.2. Lane-keeping performance

Figure 4.1.5 shows results of the absolute lateral errors for every participant, and of all participants averaged. Differences between conditions are evident in the lateral position while performing a secondary task (i.e., up to about 20 s following each magenta line, cf. Figure 4.1.4). In the Backup condition, the absolute lateral error drops to near-zero after participants were notified to perform the secondary task. In the Manual condition, however, the absolute lateral error increases with evidently higher peak values compared to periods without the secondary task. During the Forced condition, the absolute lateral error is near-zero before the secondary task periods (i.e., when automation is on) but increases substantially when the automation is disengaged.

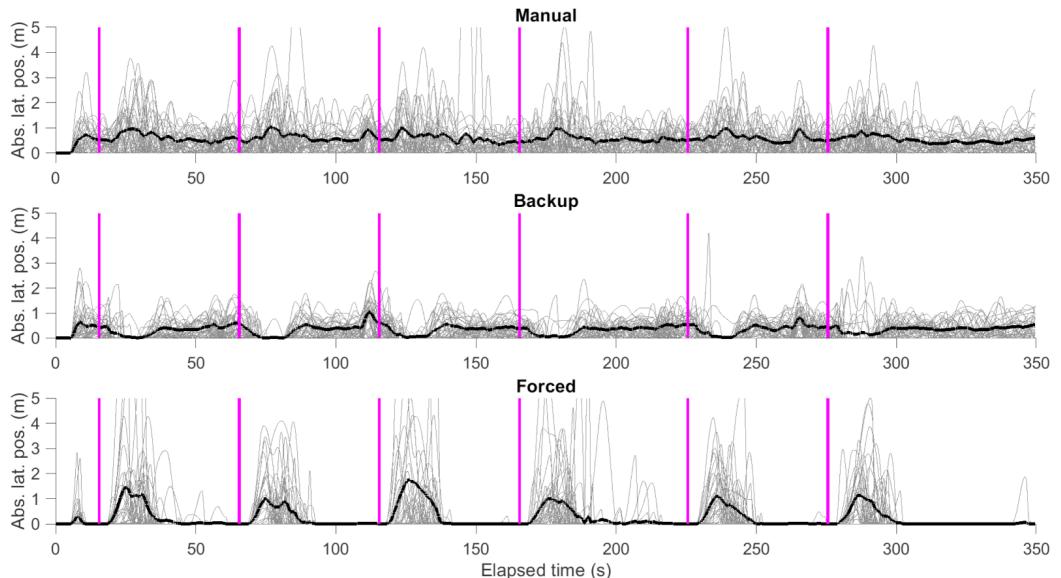


Figure 4.1.5: Absolute lateral position as a function of elapsed time. The magenta vertical lines represent the secondary task initiation beeps. The results of individual participants ($N = 31$, in each condition) are shown in gray. The mean of participants is shown in black.

Figure 4.1.6 shows the results for the three lane-keeping performance measures. Concerning the first measure (meanALE), the Backup condition yielded better lane-keeping performance than the Manual condition. Specifically, the meanALE of the Manual condition (Med = 0.54 m, IQR = 0.25 m) was higher than for the Backup condition (Med = 0.33 m, IQR = 0.12 m), $Z = 4.62$, $r = .83$, $p < .001$. The meanALE of the Forced condition (Med = 0.18 m, IQR = 0.20 m) was significantly lower than that of both the Manual condition ($Z = 4.78$, $r = .86$, $p < .001$) and the Backup condition ($Z = 3.02$, $r = .54$, $p = .003$).

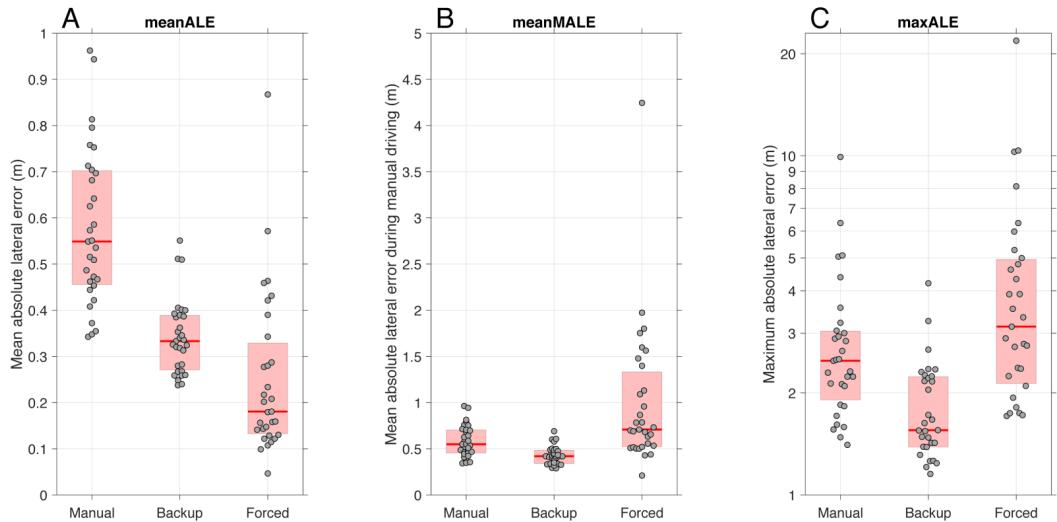


Figure 4.1.6: (A) The mean absolute lateral position (meanALE), (B) the mean absolute lateral position during manual driving (meanMALE), and (C) the maximum absolute lateral position (maxALE). maxALE is presented on a logarithmic scale. For each box, thick red the horizontal line is the median, and the edges of the box are the 25th and 75th percentiles. The markers represent scores for individual participants, with a horizontal offset to prevent overlap.

Concerning the second measure (meanMALE), which compares only the portions of manual driving, the median value for the Backup condition was 0.42 m (IQR = 0.14 m), which was significantly lower than the Manual condition (Med = 0.55 m, IQR = 0.25 m), $Z = 3.94$, $r = .71$, $p < .001$. The Forced condition yielded a significantly higher meanMALE (median = 0.71 m, IQR = 0.81 m) than the Manual condition ($Z = 3.88$, $r = .70$, $p < .001$) and the Backup condition ($Z = 4.66$, $r = .84$, $p < .001$). In summary, average lane positioning during periods of manual control with adaptive transitions of control was improved in the Backup condition compared to full manual control and was worsened in the Forced condition.

Finally, concerning the third measure (maxALE), the Manual condition yielded poorer performance (Med = 2.49 m, IQR = 1.74 m) than the Backup condition (Med = 1.67 m, IQR = 0.70 m), $Z = 3.51$, $r = .63$, $p < .001$. Furthermore, the maxALE of the Forced condition (Med = 3.14 m, IQR = 2.67) was significantly higher than that of the Backup condition, $Z = 4.23$, $r = .76$, $p < .001$, whereas the difference in maxALE between the Forced and Manual conditions was not statistically significant, $Z = 2.02$, $r = .36$, $p = .044$. In summary, maximum lane deviations were lowest in the Backup condition.

3.3. Driver attention and secondary task performance

Figure 4.1.7 shows the percentage of participants glancing at the monitor as a function of elapsed time for the three conditions. It can be seen that participants in the Backup condition were more likely to look away from the monitor (between about 3 and 12 s after the task initiation beep) than participants in the other two conditions. Participants apparently used the available backup to concentrate on the secondary task, whereas in the Manual and Forced conditions, participants had to periodically check the road to keep the vehicle in the lane. This was also reflected in the average number of points earned across the six sessions, with median values of 8.50, 9.00, and 8.67 on the scale from 0 to 10, for the Manual, Backup, and Forced conditions, respectively (Figure 4.1.8). The score for Backup was significantly higher than for the Manual ($Z = 3.02, r = .54, p = .003$) and Forced condition ($Z = 3.23, r = .58, p = .001$). The difference between the Manual and Forced conditions was not significant ($Z = 0.58, r = .10, p = .562$).

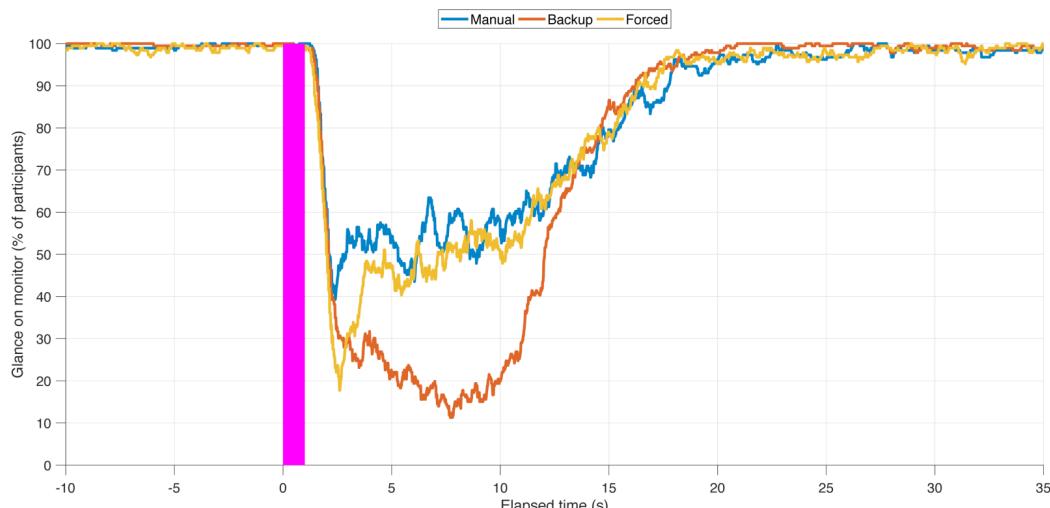


Figure 4.1.7: The percentage of participants glancing at the monitor as a function of elapsed time. Filtering with an interval of 0.25 s was applied, and the data for the six secondary tasks were averaged. Missing data (e.g., the eye tracker not tracking the eyes because the participant is performing a blink or performing the secondary task) were coded as an off-monitor glance. The thick magenta vertical line represents the secondary task initiation beep.

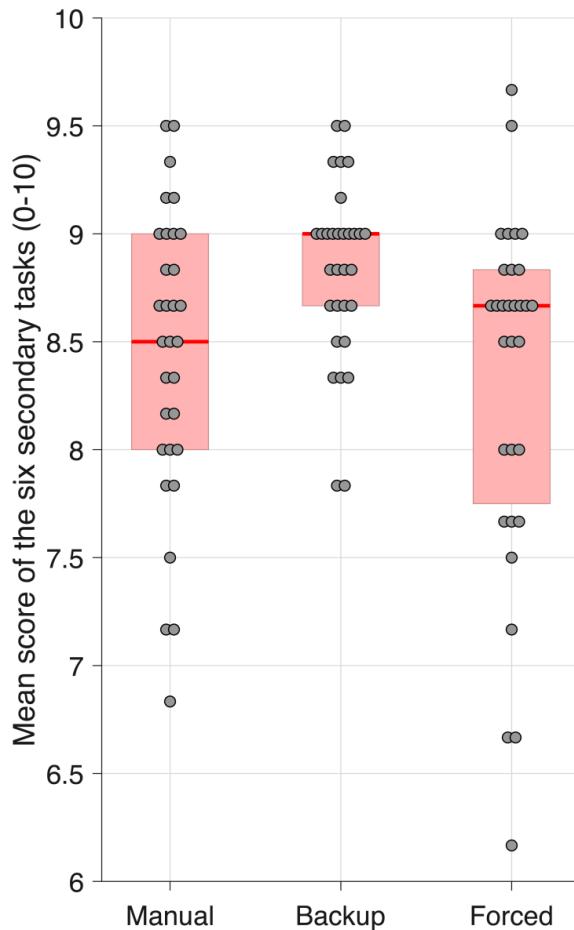


Figure 4.1.8: Points scored on the secondary task. The participant's score is the average of six tasks. For each box, thick red the horizontal line is the median, and the edges of the box are the 25th and 75th percentiles. The markers represent scores for individual participants, with a horizontal offset to prevent overlap.

3.4. Self-reported workload

The results of the NASA-TLX questionnaires per item are shown in Figure 4.1.9. Generally, the Backup condition yielded lower workload ratings than the Manual and Forced conditions for each of the six items. Regarding composite workload (i.e., the mean across the six items), the medians across participants for Manual, Backup, and Forced were 46.7%, 31.7%, and 46.7%, respectively. The composite workload of the Backup condition was significantly lower than both the Manual condition ($Z = 3.98, r = .71, p < .001$) and the Forced condition ($Z = 3.95, r = .71, p < .001$). The difference between the Forced and Manual conditions was not statistically significant, $Z = 1.09, r = .20, p = .275$.

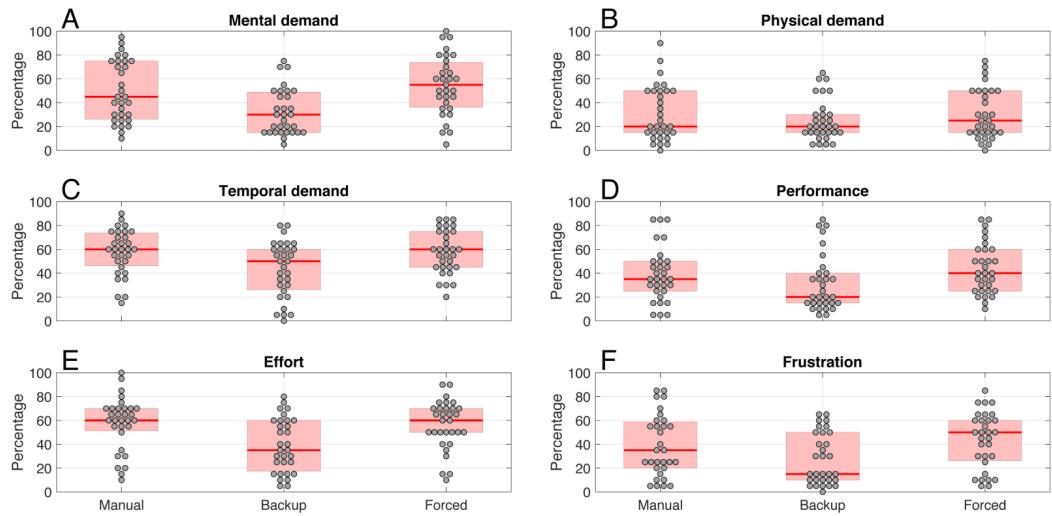


Figure 4.1.9: Results for the six items of the self-reported workload (NASA-TLX) per condition. For each box, thick red the horizontal line is the median, and the edges of the box are the 25th and 75th percentiles. The markers represent scores for individual participants, with a horizontal offset to prevent overlap. The items were answered on scale ranging from 0 = 'very low' ('perfect' for the performance item) to 100 = 'very high' ('failure' for the performance item).

3.5. Self-reported driver acceptance

The results of the acceptance scale per item are shown in Figure 4.1.10. Participants reported significantly higher acceptance scores on all items ($p < .001$) for the Backup condition as compared to the Forced condition, except for the *Raising alertness –Sleep-inducing* item.

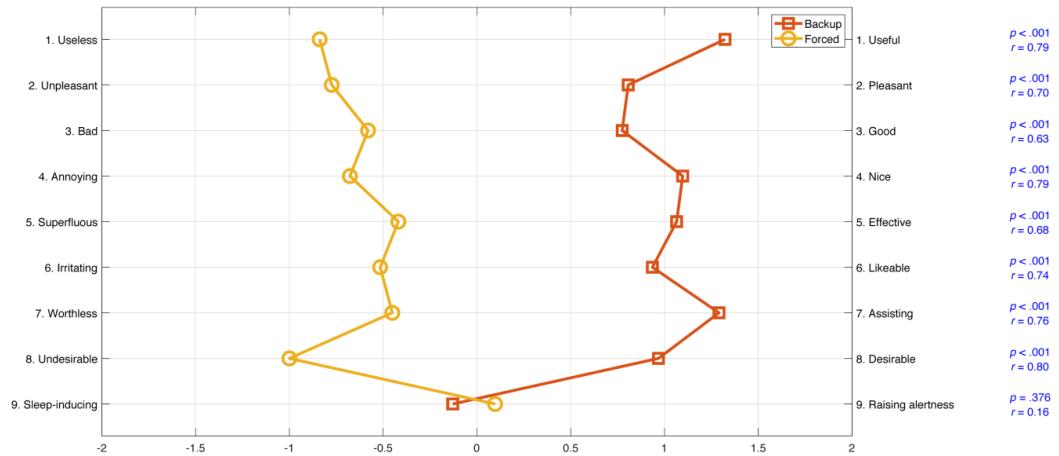


Figure 4.1.10: Mean ratings on the acceptance scale for each of the nine items. The semantic differential scale runs from -2 to 2. The figure also shows the p values and effect sizes of a Wilcoxon signed-rank test comparing the Backup condition with the Forced condition per item.

At the end of the experiment, each participant completed a form where they were asked which session they liked most. Out of the 31 participants, 22 (71%) selected the Backup condition as their preferred condition, eight (26%) selected the Manual condition, and one (3%) participant selected

the Forced condition. A final optional comments section was provided through which 13 participants provided responses (see <https://doi.org/10.4121/uuid:49d87edc-07a6-4f07-a5e6-0b699705881b>). Three participants reported that they would prefer to change control manually. Furthermore, three participants commented on the automation status bar, which was perceived as annoying, useless, and/or interfering with the working of the systems.

4. Discussion

This research aimed to design and investigate a distraction-mitigation system that automatically invoked a control transition based on distraction measurements and to see how it would affect performance, workload, and acceptance. Triggers were designed and implemented under two essentially opposite approaches to eye-based adaptive driving automation to examine different directional consequences upon detection of distraction: a transition from manual to automated control vs. a transition from automated to manual control.

4.1. Lane-keeping performance

4.1.1. Backup vs. manual

Lane-keeping performance was assessed via three complementary measures: meanALE, meanMALE, and maxALE. All three performance indices were significantly better for the Backup condition compared to the Manual condition. The substantially lower meanALE and maxALE are a direct result of the secondary task that induced visual distraction and triggered the lane centring driving automation. For the meanMALE measure, the enhanced lateral driving performance during periods of manual driving could be explained by a 'staging' benefit in the sense that the automated agent positioned the car in the center of the lane before returning manual control to the driver. However, it could also be because drivers felt more at ease and confident during manual driving, knowing that they had an automated driving agent to support them.

4.1.2. Forced vs. manual

Regarding the Forced condition, improved lane-keeping performance compared to the Manual condition was found only for the overall performance of meanALE, whereas a performance detriment was found for meanMALE. The superior meanALE of the Forced condition can be explained because automated steering was enabled for the majority (76%) of the driving time.

The fact that the Forced condition yielded lower meanALE but higher meanMALE than the Manual condition indicates that a trade-off exists between automation use (i.e., more automation is better, as automation yields zero lateral error, thereby contributing to low meanALE) and automation reliability (i.e., if drivers are required to take over, as in the Forced condition, large performance errors can result). This cost-benefit trade-off resembles the lumberjack effect, where automation has benefit for routine system performance, but a negative impact when the human has to take over (Onnasch et al., 2014).

Whether the driver is constantly in control of steering or whether he or she is occasionally forced to take control when looking away from the forward road, the maxALE did not obtain significant difference. An explanation for the observably large maxALE during the Forced condition could be that the steering wheel was not always centered during an automation-to-manual transition (see <https://doi.org/10.4121/>

[uuid:49d87edc-07a6-4f07-a5e6-0b699705881b](#) for steering angle results). Whether or not a steering wheel should be decoupled during automated driving has been a topic of debate (Kerschbaum, Lorenz & Bengler, 2014). Our results suggest that a decoupled steering wheel is associated with increased lateral positioning variability if the decoupled steering wheel is not centered at the moment of transferring control back to the human driver.

4.2. Workload

The Backup condition received the lowest self-reported workload ratings. During the Forced condition, drivers were monitoring what the automated pilot was doing, until they were forced back into control during periods of visual distraction. In other words, drivers initially experienced a state of low task demands and were forced into high task demands. This was not the case for the Backup condition, where adaptive automation was applied to help the human when task demands increased.

It should be noted that low workload ratings are not necessarily desirable, because low workload, or ‘underload’, may be associated with fatigue and loss of vigilance (Hancock & Parasuraman, 1992; Young & Stanton, 2007). Parasuraman (2003) argued that ‘clumsy automation’ can be an issue, whereby (adaptive) automation inadvertently adds workload (e.g., via new task demands like supervising or re-programming the automation) during already high periods of demand and do little to regulate workload during low periods of demands (i.e., during routine operation of the automation). In the end, an ‘optimal’ and balanced workload level should be aimed for. Within the current study, it is believed that the Backup condition supported such a balanced workload during driving because the presently 100% reliable automated steering did not require attention from the driver when it became active and it counteracted degraded lateral performance that would otherwise occur due to the uptake of the non-driving task.

4.3. System acceptance

The Backup condition was rated more favorably than the Forced condition on nearly all items, except for the *Raising Alertness–Sleep-inducing item* where results were mixed and inconclusive. The task demands in the Forced condition provide an explanation of its negative acceptance ratings. Before the transition of control, participants experienced simultaneous demands to both monitor the automated driving and undertake the secondary task. Likewise, after the transition of control, participants were still involved in the secondary task when manual control was returned to them.

The results in Figure 4.1.7 showed that participants looked away in higher proportions in the Backup condition than in the Manual condition. This suggests that the participants trusted that the automation would assume control and were more inclined to keep their focus on the secondary task. One of the intended goals of the Forced condition was to prevent drivers from misusing driving automation that requires their active oversight and mental involvement (SAE Level 2 automation). However, the Forced condition appeared to show slightly more off-road glancing than the Manual condition. This is contrary to what was intended and expected with the Forced condition design as it was meant to return driver attention to the road. Apparently, in manual driving, participants are more conservative with their off-road glances than when automation is present (whether backup or forced). This may be because, in the former, there is one driving agent in the system whereas in the latter there are two driving agents.

When asked to complete a form at the end of the experiment, a majority of participants (22 of 31) preferred the Backup condition, which supports the results from the acceptance scale. These preferences add to the promise of the Backup condition in real-world applications. However, these preferences might also be because the automation lasts for as long as the driver keeps the eyes off the road, and so allows for unrestricted secondary task engagement. A driving simulator study by Jamson et al. (2013) found results which suggested that “drivers are happy to forgo their supervisory responsibilities in preference of a more entertaining highly-automated drive”, whereas a test-track study by Llaneras, Salinger & Green (2013) showed that, when using reliable automation, drivers are likely to increase the frequency of secondary task interactions and engage in tasks that cause extended glances away from the road. In a review by De Winter et al. (2014), it was found that relative to manual driving (100%), highly automated driving resulted in 261% of the number of tasks completed on an in-vehicle display. These findings suggest that the Backup condition might be preferred because it has the potential (whether intended or not by designers) to allow for increased end-user involvement in non-driving tasks.

4.4. Limitations and generalizability

4.4.1. Driving task simplicity

The track that the participants experienced was designed to be short-lasting (350 s per drive) and easy: no obstacles, other road users, or emergency situations were implemented. Furthermore, participants were instructed to keep the center of the lane and there was also no active penalty involved with an unintended lane crossing or large lateral position errors, and there was a reward for performing the secondary task well (in the form of a post-trial feedback score which was determined by the experimenter while the participant was performing the task). These factors may have caused participants to focus on the secondary task more than they would do in real life. Future research should establish how the adaptive automation would function in more naturalistic driving conditions.

4.4.2. Eye-tracker capabilities

The eye tracker sometimes lost sight of the eyes of the driver and thus reported a null value for the gaze direction. The tracker appeared to have more difficulty with some drivers when compared to others. For our research we used a simple binary criterion to assess visual distraction: does the participant look at the monitor or not? This criterion was combined with a filter of 1.5 and 4.5 s interval (see ‘Experimental conditions’), which accounted for short data gaps due to e.g., blinking. Based on the results in Figure 4.1.4, sensitivity of the on-monitor attention algorithms must have been high, as the percentage of participants for whom the automation was ‘on’ in the Backup condition was mostly zero when participants were supposed to look at the road (i.e., in between the secondary task periods). There were a few participants for whom the automation turned on during such periods in the Backup condition; we were unable to determine whether these were due to data losses of the eye-tracker or whether participants were actually looking away from the screen (e.g., exploring whether the Backup system was working properly). Specificity must also be high because it would be unlikely for the eye tracker to measure that a participant is looking at the monitor (which subtends a relatively small angular area in the participant’s field of view) when he/she is looking instead at the CD-player. In summary, there were a few unexpected control transitions in between the secondary task periods, but these were infrequent and probably did not have a significant influence on the performance results.

The eye tracker used during this study had to be calibrated for every participant and sometimes still had trouble discerning the correct gaze direction. If the eye tracker were to calibrate itself and become more sensitive to gaze direction and less sensitive to confounding factors such as ambient lighting, this would increase the possibilities for real-world applications. Similar conclusions were drawn by Pohl, Birk & Westervall (2007) who also performed a study on distractions leading to lane departures.

4.4.3. Realism of the steering wheel

The steering wheel that was used was smaller than an actual steering wheel and was designed without any force feedback. Some participants mentioned that this lack of force feedback was annoying. Another comment some of the participants made was that it was difficult for them to follow the instructions, which stated to completely focus on what they were doing with their hands. They were told that they were not allowed to perform any part of the secondary task blindly, to prevent a situation where they could just keep looking at the monitor and still finish the secondary task in time. This, understandably, might have felt unrealistic for the task of switching a CD in the CD-player (i.e., a task people may have sufficient practice with, and which could in principle be completed without continuous visual attention). Nonetheless, this approach was implemented to ensure that control transitions did take place and to simulate situations where long consecutive eyes-off-road periods did occur.

4.4.4. Capabilities of the distraction detection algorithm

The initiation and termination threshold criteria for automatic transitions of control between the human and the automated driving were established based on pilot studies. However, these times are not necessarily generalizable, and would have to be determined again for experiments that use a different setup. For example, some drivers kept looking back and forth between the secondary task and the primary task at a high frequency. Due to this behavior, the algorithms never counted enough samples of looking away from the monitor which prevented the system from automatically initiating a control transition. A follow-up experiment could focus on discovering recommended initiation and termination times, or perhaps even incorporate an algorithm for using variable times.

The difference between safe and unsafe glances was defined by looking at the monitor or away from the monitor, respectively. In real-world driving situations, this would have to be defined more clearly. For example, further experiments might focus on what is considered as a safety region in the visual field. Perhaps it might be better to define a gradient where looking at the road directly in front of the car is or at task-relevant objects is considered to be 100% safe, whereas looking to the sides is less safe. Using such a gradient, the amount of time after which automation engages might also be varied so that, for example, a '10% safety area' uses a shorter initiation time than an '80% safety area'. It should also be noted that no mirrors were used during this experiment. Drivers usually look at the mirrors, and an improved algorithm should not classify mirror usage as a visual distraction.

Definitions of driver distraction (see Pettitt, Burnett & Stevens, 2005) are important for reliable driver monitoring and cross-study comparisons. Driver distraction can be separately categorized as visual, auditory, biomechanical, and cognitive (Ranney et al., 2000). It should be noted that the Backup and Forced systems detected visual distraction, not other types of distraction. For example, cognitive distraction is regarded as an important contributor to crashes, yet is a concept that is

hard to define (Young, 2012). Cognitive distraction in driving (Strayer et al., 2013) has been discussed in different guises, including daydreaming (Galéra et al., 2012), mind wandering (Yanko & Spalek, 2013), looked-but-failed-to-see errors (Sabey & Staughton, 1975; Staughton & Storie, 1977; Labbett & Langham, 2006), cognitive tunneling (Reimer, 2009), attention focusing (Chapman & Underwood, 1998), loss of covert/peripheral attention via diminished functional field of view (Crundall, Underwood & Chapman, 1999), and highway hypnosis (Wertheim, 1978). We reiterate here that our Backup and Forced concepts cannot detect all forms of driver aberration: in reality, drivers may drive in an unsafe manner or crash into objects even when their eyes are on the road (Victor et al., 2018), and one should therefore not expect that the present Backup automation is a remedy to all types of driver distraction. However, given the predominant importance of visual information for driving (Sivak, 1996), the generally presumed eye-mind hypothesis where gaze direction is a strong correlate of cognitive activity (Just & Carpenter, 1980), and a substantial history of driving visual occlusion research (e.g., Senders et al., 1967; Van der Horst, 2004), adaptive automation based on visual attention alone could reasonably be expected to offer a beneficial contribution.

4.4.5. Realism of the secondary task

The secondary task of changing a CD during this study was chosen because it was assumed to involve similar visual-manual loads as a number of common and risky in-vehicle tasks (e.g., texting, reaching for a dropped object, searching within a bag or purse, handling cables of charging devices, etc.). Participants were periodically forced to perform this secondary task at pre-defined moments during driving. This might have felt unnatural to some of the drivers because normally, a driver might choose a moment during driving before he or she would start a secondary task, whereas during this study these moments were forced.

4.4.6. Mode errors and human machine interface

Because of the automatic and dynamic switching of driving task responsibility between the driver and the automated driving system, the Backup and Forced conditions could be susceptible to mode confusions, a well-known problem in human-automation interaction (e.g., Feldhütter, Segler & Bengler, 2017; Sarter & Woods, 1995). A mode confusion occurs when the driver believes that the automation is on while it is off, or vice versa (see Janssen et al., in press for a framework of mode confusions in automated driving).

In our study, the status of the automation was communicated visually to the driver by means of a status bar in the middle of the dashboard. However, because the secondary task imposed a visual distraction, it was difficult for the driver to know whether the automation had taken control or not, as predicted by the multiple resource theory (Wickens, 2002). In more complex driving tasks, where the driver performs many head movements (e.g., looking over the shoulder, looking in mirrors), the driver may be susceptible to mode confusion, as such conditions could cause the Backup automation to enable itself without the driver being aware of this.

A proper human-machine interface is essential to prevent such confusions and facilitate trust in the adaptive system. Donmez et al. (2006) found that display modality of a distraction-mitigation feedback system had a strong effect on driver acceptance and trust. Future research could be focused on how to best communicate the automation status to a visually distracted driver and whether the existence of backup automation needs to be communicated at all. For example, if the

automated driving functions are implemented in an innocuous manner (e.g., small accelerations, minor corrections, blended inputs, etc.), automation status might even be best hidden to avoid confusion or misuse. That is, perhaps the driver does not need to know that the automation exists at all or when it is functioning (cf. electronic stability control, emergency enhanced braking power, etc.).

5. Conclusions and Recommendations

In conclusion, the Backup condition shows the potential to increase safety when compared to manual driving. A system that forces manual control back upon the driver appeared to be less safe than normal manual driving and less accepted than a backup system.

The current systems were designed to be simple and will need to be tested in more realistic long-lasting studies before any definitive conclusions can be drawn about the safety implications during real-world driving, and see Kalra & Paddock (2016) for calculations indicating that hundreds of millions of kilometers need to be driven in order to prove that automated driving technology is safe. Further testing might focus on expanding the simulation and the algorithm to account for other traffic, objects, emergency situations and increase fidelity by including car mirrors, and a more realistic car interior.

Finally, we note that the Backup and Forced conditions rest on different philosophies. That is, the Backup automation is a form of background automation (Kyriakidis et al., *in press*), where automation is engaged only when the driver is measured to be distracted. The assumption here is that, even though the automated driving system may be imperfect, automation is still better than a visually impaired human. The Forced automation system is a form of foreground automation, where the automation is active for most of the time but needs a human supervisor at all times. In the Backup condition, participants could devote themselves more to the secondary task than in the Forced condition. This difference result could be interpreted as good (because a given secondary task is completed sooner) or bad (because it affords the ability to devote attention to the secondary task), depending on the context of operations.

It may take many decades of technological progress until fully automated (i.e., autonomous) driving is commercially viable (Shladover, 2016). Until that time, foreground and background automation strategies are viable candidates to be further researched developed before wide-market deployment on public roads.

Acknowledgments

The research presented in this paper was conducted within the project HFAuto – Human Factors of Automated Driving (PITN-GA-2013-605817).

References

Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, Bollen KA, Brems B, Brown L, Camerer C+62 more. 2017. Redefine statistical significance. *Nature Human Behaviour* 2:10-6

Byers JC, Bittner Jr AC, Hill SG. 1989. Traditional and raw task load index (TLX) correlations: are paired comparisons necessary? In: Mital A, ed. *Advances in industrial ergonomics and safety*. London: Taylor and Francis. Vol. I:481-485

Cadillac. 2018. Introducing Cadillac super cruise: giving you the freedom to go hands-free.

Casner SM, Hutchins EL, Norman D. 2016. The challenges of partially automated driving. *Communications of the ACM* 59:70-77

Chapman PR, Underwood G. 1998. Visual search of driving situations: danger and experience. *Perception* 27(8):951-964

Crundall D, Underwood G, Chapman P. 1999. Driving experience and the functional field of view. *Perception* 28:1075-1087

De Visser E, Parasuraman R. 2011. Adaptive aiding of human-robot teaming: effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making* 5:209-231

De Winter JCF, Happee R, Martens MH, Stanton NA. 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: a review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour* 27:196-217

Donmez B, Boyle LN, Lee JD. 2007. Safety implications of providing real-time feedback to distracted drivers. *Accident Analysis & Prevention* 39:581-590

Donmez B, Boyle LN, Lee JD, McGehee DV. 2006. Drivers' attitudes toward imperfect distraction mitigation strategies. *Transportation Research Part F: Traffic Psychology and Behaviour* 9:387-398

Feldhütter A, Segler C, Bengler K. 2017. Does shifting between conditionally and partially automated driving lead to a loss of mode awareness? In: International conference on applied human factors and ergonomics. Cham: Springer. 730-741

Galéra C, Orriols L, M'Bailara K, Laborey M, Contrand B, Ribéreau-Gayon R, Masson, F, Bakiri, S, Gabaude, C, + 5 More. 2012. Mind wandering and driving: responsibility case-control study. *British Medical Journal* 345:e8105

Hancock PA. 2007. On the process of automation transition in multitask human-machine systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 37:586-598

Hancock PA, Parasuraman R. 1992. Human factors and safety in the design of intelligent vehicle-highway systems (IVHS). *Journal of Safety Research* 23:181-198

Hart SG, Staveland LE. 1988. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N, eds. *Human mental workload*. Amsterdam: North Holland Press. Vol. 1:139-183

Horberry T, Anderson J, Regan MA, Triggs TJ, Brown J. 2006. Driver distraction: the effects of concurrent in vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention* 38:185-191

Inagaki T. 2003. Adaptive automation: sharing and trading of control. In: Hollnagel E, ed. *Handbook of cognitive task design*. Mahwah: Erlbaum. 147-169

Jamson AH, Merat N, Carsten OM, Lai FC. 2013. Behavioural changes in drivers experiencing highly automated vehicle control in varying traffic conditions. *Transportation Research Part C: Emerging Technologies* 30:116-125

Janssen CP, Boyle L, Kun A, Ju W, Chuang L. A hidden Markov framework to capture human-machine interaction in automated vehicles. *International Journal of Human Computer Interaction* In Press

Just MA, Carpenter PA. 1980. A theory of reading: from eye fixations to comprehension. *Psychological Review* 87:329-354

Kaber DB, Endsley MR. 2004. The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science* 5:113-153

Kalra N, Paddock SM. 2016. Driving to safety: how many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A* 94:182-193

Kerschbaum P, Lorenz L, Bengler K. 2014. Highly automated driving with a decoupled steering wheel. In: Proceedings of the human factors and ergonomics society annual meeting (Vol. 58, No. 1). Sage. Sage Publications. Vol. 58, (1):1686-1690

Klauer SG, Dingus TA, Neale VL, Sudweeks JD, Ramsey DJ. 2006. The impact of driver inattention on near

crash/crash risk: an analysis using the 100-Car Naturalistic Driving Study data. Tech Rep. No. DOT HS 810 594. National Highway Traffic Safety Administration, Washington, D.C.

Kyriakidis M, De Winter JCF, Stanton N, Bellet T, Van Arem B, Brookhuis K, Martens MH, Bengler K, Andersson J, Merat N+4 more. A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science* In Press

Labbett S, Langham M. 2006. What do drivers do at junctions? In: Proceedings of the 71st Royal Society for the Prevention of Accidents (RoSPA) international road safety conference. Blackpool. 1-13

Llaneras RE, Salinger J, Green CA. 2013. Human factors issues associated with limited ability autonomous driving systems: drivers' allocation of visual attention to the forward roadway. In: Proceedings of the 7th international driving symposium on human factors in driver assessment, training, and vehicle design. Bolton Landing. 92-98

Liang Y, Lee JD, Horrey WJ. 2014. A looming crisis: the distribution of off-road glance duration in moments leading up to crashes/near-crashes in naturalistic driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 58:2102-2106

Matthews G. 2016. Multidimensional profiling of task stress states for human factors: a brief review. *Human Factors* 58:801-813

National Transportation Safety Board (NTSB). 2017. Collision between a car operating with automated vehicle control systems and a tractor-semitrailer truck.

Onnasch L, Wickens CD, Li H, Manzey D. 2014. Human performance consequences of stages and levels of automation: an integrated meta-analysis. *Human Factors* 56:476-488

Parasuraman R. 2000. Designing automation for human use: empirical studies and quantitative models. *Ergonomics* 43:931-951

Parasuraman R. 2003. Adaptive automation matched to human mental workload. In: Hockey GRJ, Gaillard, AWK, Burov O, eds. *Operator functional state: the assessment and prediction of human performance degradation in complex tasks*. Amsterdam: IOS Press.

Parasuraman R, Bahri T, Deaton AE, Morrison JG, Barnes M. 1992. Theory and design of adaptive automation in aviation systems. Progress report No. NAWCADWAR-92033-60. Warminster: Naval Air Warfare Center, Aircraft Division

Parasuraman R, Riley V. 1997. Humans and automation: use, misuse, disuse, abuse. *Human Factors* 39:230-253

Peng Y, Boyle LN, Hallmark SL. 2013. Driver's lane keeping ability with eyes off road: insights from a naturalistic study. *Accident Analysis & Prevention* 50:628-634

Pettitt M, Burnett G, Stevens A. 2005. Defining driver distraction. In: Proceedings of the 12th intelligent transportation systems world congress. San Francisco.

Pohl J, Birk W, Westervall L. 2007. A driver-distraction-based lane-keeping assistance system. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering* 221:541-552

Ranney T, Mazzae E, Garrott R, Goodman M. 2000. NHTSA driver distraction research: past, present and future. Technical report. National Highway Traffic Safety Administration, Washington, D.C.

Reimer B. 2009. Impact of cognitive task complexity on drivers' visual tunneling. *Transportation Research Record* 2138:13-19

Ryu JB, Sihn YK, Yu SB. 2013. Assessment of risky driving caused by the amount of time focused on an in-vehicle display system. *International Journal of Automotive Technology* 14:259-264

Sabey BE, Staughton GC. 1975. Interacting roles of road environment, vehicle, and road user in accidents. In: 5th international conference of the international association for accident traffic medicine. London. Transport and Road Research Laboratory (TRRL), Crowthorne, Berkshire.

SAE International. 2016. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems.

Sarter NB, Woods DD. 1995. How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors* 37:5-19

Scerbo MW, Freeman FG, Mikulka PJ, Parasuraman R, Di Nocero PJ, Prinzel LJ. 2001. The efficacy of psychophysiological measures for implementing adaptive technology (NASA TP-2001-211018). NASA Langley, Hampton

Senders JW, Kristofferson AB, Levison WH, Dietrich CW, Ward JL. 1967. The attentional demand of automobile driving. *Highway Research Record* 195:15-33

Sheridan TB, Parasuraman R. 2005. Human-automation interaction. *Reviews of Human Factors and Ergonomics* 1:89-129

Shladover S. 2016. The truth about self-driving cars: they are coming, but not the way you may have been led to think. *Scientific American* 314:52-57

Sivak M. 1996. The information that drivers use: is it indeed 90% visual? *Perception* 25:1081-1089

Staughton GC, Storie VJ. 1977. Methodology of an in-depth accident investigation. Survey report no.

672. Transport and Road Research Laboratory (TRRL), Crowthorne, Berkshire

Strayer DL, Cooper JM, Turrill J, Coleman J, Medeiros-Ward N, Biondi F. 2013. *Measuring cognitive distraction in the automobile*. Washington, D.C.: AAA Foundation for Traffic Safety.

Van der Horst R. 2004. Occlusion as a measure for visual workload: an overview of TNO occlusion research in car driving. *Applied Ergonomics* 35:189-196

Van der Laan JD, Heino A, De Waard D. 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies* 5:1-10

Victor TW, Tivesten E, Gustavsson P, Johansson J, Sangberg F, Ljung Aust M. 2018. Automation expectation mismatch: incorrect prediction despite eyes on threat and hands on wheel. *Human Factors* Epub ahead of print Aug 10 2018

Wertheim AH. 1978. Explaining highway hypnosis: experimental evidence for the role of eye movements. *Accident Analysis & Prevention* 10:111-129

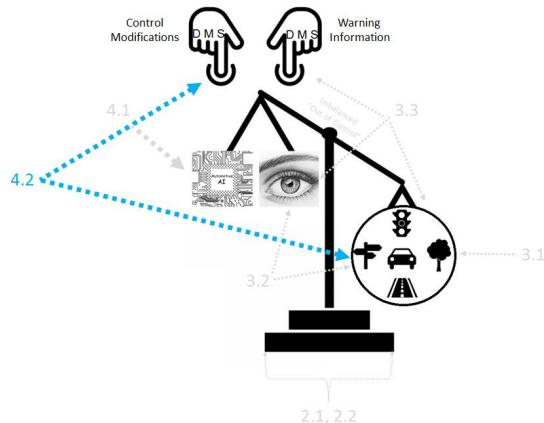
Wickens CD. 2002. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science* 3:159-177

Yanki MR, Spalek TM. 2013. Driving with the wandering mind: the effect that mind-wandering has on driving performance. *Human Factors* 56:260-269

Young MS, Stanton NA. 2007. What's skill got to do with it? Vehicle automation and driver mental workload. *Ergonomics* 50:1324-1339

Young R. 2012. Cognitive distraction while driving: a critical review of definitions and prevalence in crashes. *SAE International Journal of Passenger Cars-Electronic and Electrical Systems* 5:326-342

Chap. 4.2) Redesigning Today's Driving Automation Towards Adaptive Backup Control with Situated and Implicit Interfaces



In regards to the overall thesis big picture, this driving simulator experiment serves to provide initial explorations of different variations on adaptive backup driving automation (i.e., which triggers on/off upon detection of visual distraction/attention). Eye tracking was used here on an applications basis but not investigated as a primary research factor of interest. Firstly, problems were confirmed for supervisors of driving automation where more non-response errors were made to unexpected hazards than by those with full conventional/manual control. Such results substantiate a motivating interest in alternative functional allocations of driving automation than those conceptually similar to what is presently being released in the automotive market. Instead adaptive backup driving automation was seen to improve lateral control compared to manual driving (consistent as with Chap. 4.1) and with lower levels of visual distraction and fewer non-response errors compared to supervised automated driving. Using a scene-tied implementation of inattention (i.e. situated) effectively reduced the number of unnecessary alerts (i.e., without safety impact) compared to a condition where distraction was based only on looking away. Furthermore, to mitigate potential over-reliance and automation misuse (e.g., becoming distracted because you expect the automation to back you up), the status display of backup automation was removed without any negative impact on any of the present measures of safety, efficiency, performance or acceptance.

Adapted from:

Cabral, C.D.D., Stapel, J.C.J., Happee, R. & de Winter, J.C.F. (under review). Redesigning today's driving automation towards adaptive backup control with situated and implicit interfaces.

Abstract

Objective: We investigated adaptive backup designs for distracted drivers via a driver monitoring system (DMS). **Background:** Combined lateral/longitudinal driving automation backup may be an effective redesign of roles compared to assumption of human supervision of continuous automation. However, such backup control concepts pose complications: distrust of distraction assessment and/or misuse via over-reliance. **Methods:** 91 participants were assigned between-subjects to conditions of supervised automated driving and conventional driving with different forms of DMS-based adaptive backup control. We compared supervision with and without a hand-on-steering-wheel requirement, an 'eyes-only' DMS detecting visual distraction against an 'eyes-plus-situation' DMS requiring the additional presence of a course/collision conflict, and an 'explicit' backup providing display of automation status against an '*implicit*' backup without notification or driver awareness of the automation. All participants performed an NDRT (visual N-back) for the entire driving trial. **Results:** Automated driving increased visual distraction and non-responses to hazards compared to backup and conventional driving. A hand-on-the-wheel requirement improved response generation compared to no-hands-on-the-wheel. Across an entire driving trial, the backup improved lateral performance compared to conventional driving. Without negatively impacting safety, the eyes-plus-situation DMS reduced amounts of unnecessary automated control compared to the eyes-only DMS conditions. Eyes-only assessment produced low satisfaction ratings, whereas eyes-plus-situation satisfaction was on par with automated driving. Removing indication of driving automation made no appreciable difference. **Conclusions:** We evidenced the preliminary feasibility of driving automation to serve as a situated and implicit backup safety system. **Application:** Redesigns of driving automation with eye-based DMS can enable adaptive control benefits.

1. Introduction

1.1. Fatal firsts, conflicting expectations, and potential re-starts in the race to self-driving

The present experiment explores the human factors problem of expecting people to effectively supervise iterative levels of automated/autonomous vehicles (AVs) (e.g., SAE 2018) with several solutions. Given the controversial nature of anticipated AVs benefits (e.g., De Winter, in press; Bhuiyan, 2018), we take it upon ourselves to not only motivate and test our proposed solutions (situated, implicit, adaptive backup driving control), but to first motivate and test our problem expectations (inattention in human supervision of driving automation) as well as others' status quo solutions (keeping a hand on the steering wheel while supervising). So first, brief clarification of problems with the current way forward is needed and substantiated as motivation of the present counter-position detour.

News reports of the first people killed in AV crashes illustrate new levels of inattention risks while driving (Fung, 2017; Coppola & Frank, 2018). Because general AI technology to replace human driver flexibility is not yet proven despite contrary public opinion (e.g., Euro NCAP, 2018 highlights '*stark contrasts*'), it is likely that the public will become too complacent in their supervision over AVs. Furthermore, instructions to monitor AV technology are inconsistent with observed would-be consumer behaviors (Carsten et al., 2012; Jamson et al., 2013; Large et al., 2017) and value proposition/preference (Cyganski et al., 2014; Bertoncello & Wee, 2015). Even if people wanted to supervise driving automation, many decades of human factors research, from Mackworth (1950) to Hancock (2017), have suggested risks when humans monitor automated (e.g., monotonous, self-regulating, removed, etc.) processes over extended periods of mostly successful operation. Such risks have recently been substantiated by reviews specific to the driving domain (e.g., Cabrall et al., 2016a; Goncalves et al., 2017). A potential reason why ironies of automation may expand rather than eliminate problems with human operators (e.g., Strauch, 2017), is that humans are prone to unconscious switching/trading of attention rather than the even-handed attentional sharing desired for supervisors of automation. Instead, cost-free multi-tasking has been generally ousted as a '*myth*' (Loukopoulos et al., 2009; Rosen, 2008).

For human-machine teaming, there can be more rational "first-steps" than full-time driving automation that must only assume human supervisory oversight. Driver monitor systems (DMS) can warn against supervisory inattention and/or trigger transitions of control (ToC) in new adaptive function allocation designs (e.g., Petermeijer, et al., 2015; Cabrall et al., 2018a). Generally, human errors in driving (e.g., the often repeated over 90% of fatal accidents statistic from NHTSA, 2008) should be recognized as exception cases (because accidents, and fatal ones at that, are by definition exception cases already) and to thus motivate more targeted solutions (i.e., at periodic events of degraded human driver attention) rather than full system re-hauls of unknown consequences and extra risks. The present introduction proposes a reversal of AV technology away from continuous operation and towards DMS-triggered adaptive backup. In the remainder of the paper, we extend previous research with experimental comparisons of various interface designs for how that adaptive backup system might be conceived. We compare regular eye-tracking-based DMS assessments of distraction ('eyes-only') to scene-tied DMS ('eyes-plus-situation') taking driving conditions into account. We compare '*explicit*' DMS informing drivers when backup is activated to '*implicit*' DMS without notification or driver knowledge of the automated system.

1.2. Driver monitor system (DMS) solutions

The U.S. National Transportation Safety Board (NTSB) issued new safety recommendations on September 12, 2017 (NTSB, 2017) for manufacturers to '*develop applications to more effectively sense the driver's level of engagement and alert the driver when engagement is lacking while automated vehicle control systems are in use*'.

1.2.1. Hand placement

As a basic first form of DMS in SAE Level 2 AVs, many manufacturers require the driver to maintain hand contact with the wheel (Audi, BMW, Mercedes, Tesla, and Volvo). From voluntary safety self-assessments collected by NHTSA (2018), multiple '*fully*' autonomous driving vehicles (Apple, Ford, GM, and Uber) can also be seen to currently require safety-driver hand placement on/near the wheel during on-road test/development. Beyond faster responses from having closer positioning, hand-on-wheel placement may yield risk detection and memory benefits underlying the successful generation of a response. Positioning of hands-on-the-wheel has been associated with risk perception rather than only fatigue or personal style preferences (Walton & Thomas, 2005). In physical rehearsal of movements (e.g., sports, dance, etc.), memory is tightly coupled to motor processes: '*Motor practice is associated with the formation of elementary motor memories*' (Stefan et al., 2008). The primary motor cortex has been shown to hold short-term representations of recently practiced movements with encoded kinematic details (Classen et al., 1998, 1999; Butefisch et al., 2000).

1.2.2. Adaptive backup control

Previous human factors research has suggested an industrial self-affliction of vigilance problems where humans must supervise automation: Hancock (2013, 2017) described the problem as '*iatrogenic*' and Parasuraman and Riley (1997) called it an '*abuse*' by creators of automation. Thus, with more effectively designed DMS, it is worth considering alternatives to the controversial function allocation that seeks to recast human drivers into supervisors of full-time automated driving. Given that the majority of human driving is successful and safe, a more rational step would be to support periods of degraded human driver attention in a selective manner. The general notion for the reversal of continuous driving automation implementations towards event-driven backup is supported by prior human factors research that addresses degraded human vigilance in supervision of automation by use of shorter durations of supervision and schemes of adaptive control (e.g., Parasuraman & Wickens, 2008; Sheridan & Parasuraman, 2005; Parasuraman et al., 1996; Scallen et al., 1995). In particular for driving, Petermeijer et al. (2015) found benefits of event-driven backup (bandwidth feedback) to avoid negative aftereffects compared to continuous shared-control automation when automation is unexpectedly removed. Furthermore, a driving simulator study of Cabrall et al. (2018a) found that, in the presence of a distraction activity, a backup automation system performed the best (decreased lateral errors, lower self-reported workload, and higher levels of acceptance) in comparison with conventional driving control and a full-time driving automation system that automatically disengaged itself upon detecting driver distraction (e.g., a concept consistent with current forms of on-market systems).

The simulation visuals of Cabrall et al. (2018) was minimalistic (i.e., only road and grass) and hence the present experiment aims to replicate adaptive backup control benefits with increased environmental complexities and to explore further system design opportunities. Monotonous environments may aggravate inattention issues in supervision over automation: with less in the

driving environment (especially compared to real life), participants might be inclined towards greater amounts of non-driving related task (NDRT) engagement. Next, we describe additional functional and interface design considerations to avoid human-automation interaction trust issues that might likely arise within such a concept of adaptive distraction-induced backup driving automation.

1.2.3. Situated assessments of distraction

An assessment of driver distraction is expected to be insufficient by looking only at the driver, as such approach may yield false alarms and 'cry-wolf' effects. Distraction assessment should not be limited to an internal human-centric focus but could extend its perspective by looking outward of the vehicle for relevant contextual considerations. Accordingly, we propose a more conservative implementation strategy for classifying distraction (called '*situated*' adaptive automation). This approach is not yet common in the automotive AV market but is not without precedent (see Hof, 2016; Simonite, 2017). In addition to diversion of attention away from driving, the driver distraction definition of Hedlund et al. (2006) included resource competition and increases in risk, and highlighted their implication that '*distractions are affected by driving conditions and situations*'. Beyond interrogating if a driver is looking away from the road to an NDRT, a DMS might ask if the driver is looking away *too much given the present circumstances* (cf., Minimum Required Situation Awareness, Kircher & Ahlstrom, 2017). This paper extends theoretical recommendations with an applied research investigation for benefits and operational feasibility of situated assessments in a distraction-induced driving automation backup concept. Immediate lateral and longitudinal control demands form the inner-core of descriptive hierarchical driving models (e.g., Merat et al., 2019; Michon, 1978, 1985), and so seem a reasonable level to implement practical situated assessments of attention.

1.2.4. Implicit backup operations

If people believe the system will back them up, they may allow themselves to become distracted more often with expectation for backup from the automation (i.e., misuse through over-reliance). While the notion of appropriate feedback has been a mainstay constituent of good human factors design (e.g., Norman, 1990) and for advanced driver assistance systems (Seppelt & Lee, 2007), it does not necessarily imply that feedback is needed for all things at all times. An avenue for reducing operator over-reliance on SAE Level 2 driving automation might be to make its operation less apparent (i.e., '*implicit*' adaptive automation) rather than providing '*explicit*' information of system existence/activation. Reasonably, it is harder to misuse something (e.g., Parasuraman & Riley, 1997) that you do not know is there. Furthermore, explicit DMS information during a period of detected operator distraction may increase workload and unwanted visual behaviors especially if the HMI is confusing or unwanted/un-trusted. Jaguar Land Rover's head of safety, Phil Glyn-Davies, has proposed in Bird (2018) that '*the best active safety system is one where you're not even aware of its presence*'.

1.3. Research questions and aim

In summary, background evidence suggests healthy skepticism for the capabilities of current on-market AVs and of human supervisors of continuous driving automation. Tempting new opportunities for increased levels of distraction (e.g., with highly engaging/demanding NDRTs) are likely and can have fatal consequences. We assume distraction to be dangerous when it reaches levels that degrade safe vehicular control such as an ability to respond to hazards and to stay within

a target lane of travel. Thus, the present paper seeks to assess joint system outcomes in terms of human behavior and vehicle performance when DMS and driving automation components are given a different human system integration (i.e., via various designs of event-driven periodic backup support).

SAE Level 2 AV technology (simultaneous automatic lateral and longitudinal control) as exists today might be re-branded and re-implemented more towards a safety rather than a convenience feature—in other words, away from '*automation always/mostly replaces the human driver*' and towards '*automation backs up the human driver as needed*'. In order to do so, new functional and interface design considerations are needed and worth exploring. The present paper addresses 5 related research questions (RQ) embedded within a single experiment.

- (1) RQ1 – Are drivers susceptible to dangerous levels of distraction with SAE Level 2?
- (2) RQ2 – Does placing a hand on the wheel improve driver supervision of automation?
- (3) RQ3 – Is adaptive backup a safe and acceptable alternative to continuous automated driving?
- (4) RQ4 – Can situated criteria safely reduce driver state monitoring from over-triggering?
- (5) RQ5 – Is the status of backup driving automation necessary to display to drivers?

2. Methods

2.1. Participants

The experiment was completed by 91 university students (26 female, 65 male) aged between 21 and 34 years ($M = 23.51$, $SD = 2.17$) with a majority (73%) indicating a driving frequency between a weekly and monthly basis. Overall, participants had a driving license for about four and a half years ($M = 4.48$, $SD = 2.70$). This research complied with the American Psychological Association Code of Ethics and was approved by the Human Research Ethics Committee of the TU Delft. Informed consent was obtained from each participant.

2.2. Apparatus

The driving simulation hardware consisted of the Logitech G27 USB gaming steering wheel and pedals. The software was programmed within MathWorks Simulink (2017b) model-based design environment and TASS International PreScan simulation (release version 7.4) which is '*a physics-based platform that is used in the automotive industry for the development of Advanced Driver Assistance Systems (ADAS) that are based on sensor technologies such as radar, laser/lidar, camera and GPS*' (TASS International, 2019). The simulated driving visuals were displayed on an NEC MultiSync EA 243wm monitor with a 52 cm x 33 cm viewable image at 1920 x 1200-pixel resolution that was placed approximately 65 cm from participants' eyes. A SmartEye DR120 remote eye tracker was used with its cameras concealed behind a black bar beneath the simulation display monitor. Figure 4.2.1 depicts the overall apparatus.



Figure 4.2.1. Driving simulator arrangement including a gaming steering wheel attached to the edge of the table (bottom), floor pedals (not shown), driving simulation visuals displayed on a monitor (middle), and secondary task on a laptop with an external mouse (right).

2.3. Simulated level of driving automation

A version of SAE Level 2 driving automation was implemented in the driving simulator. The software kept the vehicle in the middle of the right lane at a constant speed of 70 km/h. Additionally, the vehicle automatically reduced its speed to maintain spacing as needed behind a slower lead vehicle, and returned to the target speed of 70 km/h when that slower lead vehicle moved away from the lane of travel. The driving automation included a stipulation that the participant must monitor and correct the automated driving for any dangers/errors.

2.4. Driver monitoring system using eye-tracking

Individual MathWorks Simulink architectures were built and deployed on different computers to model the driving simulation with its automated control functions separately from the DMS. These two systems were integrated for real-time operation by use of standard UDP communication channels. The DMS was designed to function by receiving participant gaze direction and eyelid opening information as inputs from the eye tracker to assess several driver states of distraction, drowsiness, and/or cognitive overload across different time period criteria specific to each state

and towards an elevated state of '*aberrance*' (Cabral et al., 2016b). For the purposes of this experiment, the DMS was used to continuously assess whether the participant was looking at the screen or not (as with Cabral et al., 2018). Classification of visual distraction was implemented by similar mechanisms as in Cabral et al. (2018). A prior state of attentive/distracted held until the threshold was met to change that state: consequently, the participant was always classified as either being attentive or distracted at any given point in time across the full driving trial. The distraction threshold approximated a 2-second criterion which has generally been accepted from the results of widely-cited driver distraction research (Klauer et al., 2006; NHTSA, 2013) and consequently frequently corroborated (e.g., Ryu et al., 2013). For example, Rockwell (1988, p. 322) states: '*For years researchers studying car following and eye movements have found a 2 second rule, i.e., drivers are loath to go without roadway information for more than 2 seconds (and rightly so)*'. In the present study, distraction states were applied after 3 consecutive seconds of looking away from the simulation display monitor (with a reset after 4 consecutive seconds of looking forward again). It should be noted that these thresholds were intended to be half as large, to be around the same levels as suggested by previous research (Kircher & Ahlstrom, 2009; Seaman et al., 2017; Seppelt et al., 2017), but a system integration error transpired where the downscaling of the eye-tracker measurement frequency (120 Hz) as limited by the driving simulation resolution (60 Hz) was not properly accounted for in the classification algorithm.

It should also be noted that previous research suggests that exact durations of off-road glances for classifying distraction could be variable and might not actually be as problematic as is an increase in the frequency of longer duration glances (see Liang et al., 2014). For example, Rockwell (1988, p. 324) states that drivers '*will pay the price in more glances but not longer glances*'. Attentional buffers of between 2.5 and 5.5 seconds for off-road glances are suggested by results of Godthelp et al. (1984), and between 2 and 4+ seconds of on-road glances from Samuel and Fisher (2015) and Glaser et al. (2016), and even upwards of between 7 and 12 or beyond 20 seconds for establishing aspects of roadway situation awareness from Lu et al. (2017). Furthermore, our methodological error in the timing of visual distraction/attention classification should not invalidate our present results, as our results are presently analyzed in a conservative manner in terms of relative comparisons between conditions (e.g., percentage) rather than in an absolute number of seconds.

2.5. Adaptive transitions of control (automated control as "Backup")

Two types of adaptive driving automation backup were evaluated via an experimental DMS. In one case ('eyes-only'), detections of driver visual distraction directly activated automated driving control functions (i.e., lateral control via steering the vehicle to the center of the right lane, and longitudinal control by gradually slowing down). In the other case ('eyes-plus-situation'), the operating routine required both the detection of driver distraction and simultaneous course/collision conflict predictions to activate the automated control functions. In either operating case, conventional driving control (human operation of steering wheel, throttle, and brake) was re-activated when all criteria for automated driving control was no longer met.

Course/collision conflict predictions were assessed with simulated radars for road departure or collision with an object. The simulated lateral and longitudinal radars each interrogated a fixed distance ahead of the vehicle (approximately 20 and 100 meters, respectively) to determine a binary state of course/collision conflict. Assuming a traveling speed of 70 km/h (a typical speed targeted in our simulation), the look-ahead positioning of these radars represented time budgets of

approximately 1 and 5 seconds for course and collision conflicts respectively. The present conflict predictions were not yet capable of dynamically adjusting their ranges based on actual driven speed fluctuations. With only a fixed look-ahead distance, actual speeds slower/faster than 70 km/h respectively increased/decreased the time budgets, and diminished/inflated the frequency of alerting and thus also the potential for backup automated driving control.

If the automatic transition of control (ToC) status was displayed (i.e., in the automated driving and explicit backup conditions), it appeared on the right side of a virtual dashboard and read either as '*Normal Driving*' (green background) or '*Auto Backup Control*' (red background). In the implicit backup conditions, the automation status was not shown, and participants were led to believe that they were driving conventionally only (see Table 4.2.1).

Table 4.2.1. Experimental conditions ($n = 13$ per group).

Condition	Automation functionality	Task instructions
EXP1. Continuous automation – no hands https://youtu.be/hTH0bqyHKS1	Automated simultaneous longitudinal and lateral control for the entire drive.	Automated Driving: <i>No hands needed on the wheel, no feet are needed on the pedals, but you must monitor and correct the automated driving for any dangers/errors</i>
EXP2. Continuous automation – one hand https://youtu.be/7ulfPk5Do_Y	Automated simultaneous longitudinal and lateral control for the entire drive.	Automated Driving: <i>One hand needed on the wheel (<i>just to touch, not to steer</i>), no feet are needed on the pedals, but you must monitor and correct the automated driving for any dangers/errors</i>
EXP3A. Backup – eyes plus situation assessment with explicit automation status https://youtu.be/SIUPseabxwU	Backup simultaneous longitudinal and lateral control if the participant was visually distracted and the situation was deemed unsafe (detected course/collision conflict prediction).	Manual Driving: but driving automation (collision avoidance, middle of right lane) may automatically turn on and off periodically to help you (it decides when/where/how long and how much). It does this from looking at the <u>road situation</u> and at your <u>eyes</u> . The automation is not perfect, so it cannot be relied upon to do all of the driving. Automation status is shown on screen in green=off/red=on.
EXP3B. Backup – eyes only assessment with explicit automation status https://youtu.be/6BS1w5uVtHk	Backup simultaneous longitudinal and lateral control if the participant was visually distracted.	Manual Driving: but driving automation (collision avoidance, middle of right lane) may automatically turn on and off periodically to help you (it decides when/where/how long and how much). It does this from looking only at your <u>eyes</u> . The automation is not perfect, so it cannot be relied upon to do all of the driving. Automation status is shown on screen in green=off/red=on.
EXP4A. Backup – eyes plus situation assessment	Backup simultaneous longitudinal and lateral control if the participant was visually distracted.	This is a manual driving condition with eye tracking that we need to

Condition	Automation functionality	Task instructions
with implicit automation status https://youtu.be/p1ly_S4YlZ8	visually distracted and the situation was deemed unsafe (detected course/collision conflict prediction). No automation status was shown on the screen.	use as a comparison against other subjects who use automation.
EXP4B. Backup – eyes only assessment with implicit automation status https://youtu.be/SrxcPauPyXE	Backup simultaneous longitudinal and lateral control if the participant was visually distracted. No automation status was shown on the screen.	Same as 4A
EXP4C. Conventional driving – No automation https://youtu.be/gwrr5796EVI	No automation; manual driving only. No automation status was shown on the screen.	Same as 4A

2.6. N-back secondary task

Our NDRT shared the expressed motivations of automotive research from Mehler et al. (2011): '*to induce varying levels of demand so that the impact on participants can be observed*' (p. 3). We aimed to place the participant in a dual-tasking state whereby he/she would be challenged to balance engagement in an activity unrelated to driving in competition with driving activity and responsibility. Because distractions involving reaching and searching have been implicated as some of the most detrimental in naturalistic vehicular safety studies (Hickman, 2015), our specific implementation of the N-back task was in a graphical user interface (GUI) format (see Figure 4.2.2) to add visual-manual demands to conventional cognitive demands (which have previously been induced most conventionally along only the auditory channel). It was also felt that this modification of the N-back task might better resemble real-life attentional demands such as with continuous time-response critical visual-manual tasks (e.g., mobile phone instant-messaging) but in a controllable manner and with empirical research precedence. Through pilot studies, it was determined that an immediate '*zero-overlap*' response level of N-back in the GUI was sufficient to impose resource competitions on driving performance in our simulator while participants were still able to achieve near perfect scores when performing that N-back in an isolated training session. It should be noted that for translation purposes, it was easier to explain and label the task as a '*1-back*' for our non-native English speaking participants although the task was the conceptual analog of the '*0-back*' as described within Mehler et al. (2011).

Our visual N-back application is available online (Cabral, 2017). Demonstration videos by an experimenter are available from URLs in Table 4.2.1. As shown in Figure 4.2.1, the placement of the N-back GUI was to the right (about 35–45 degrees) and slightly below the dashboard of the driving simulator by a few inches and within arms-reach (e.g., in rough positional correspondence to a center-console display, although participants used a mouse to input their responses). The pacing involved an equivalently matched target display and response allotment time that randomly varied between 1 and 2.25 seconds (at 0.25-second resolution). The same scripted set of pre-randomized timings was used for every participant. The participants were informed that they would be scored on the N-back task with correct answers receiving +1 point and incorrect/missed answers receiving -1 point. Auditory feedback included a '*beep*' for a correct answer, a '*buzz*' for an incorrect answer, and silence for a missed answer. Otherwise, scores were not displayed or communicated. Participants were not told what they should prioritize, other than that they should do their best to

simultaneously balance both their driving/supervision responsibilities (see Table 4.2.1) along with the N-back task.

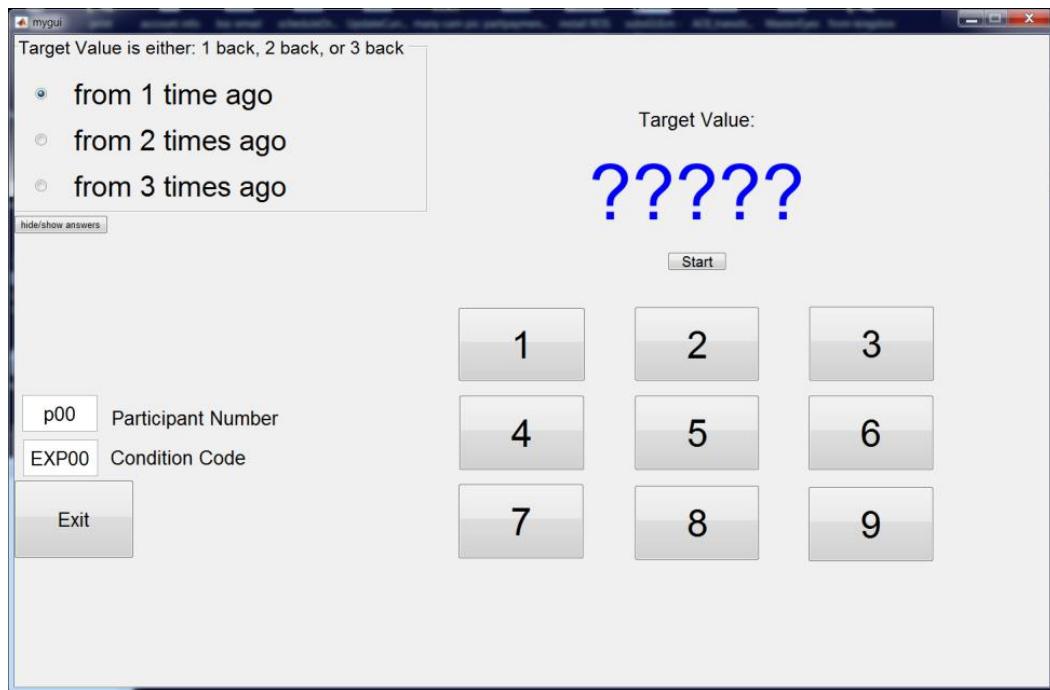


Figure 4.2.2. A modified N-back task was used as a secondary task presented via a graphical user interface (GUI).

2.7. Conditions and procedures

A between-subjects experiment was conducted. Participants were randomly allocated to one of seven experimental conditions, as shown in Table 4.2.1. Upon examination of the results, two participants were removed from the analyses. A participant from EXP3A was removed because his experimental condition was mistakenly inconsistent with his provided instructions (i.e., wrong condition). A participant from EXP4C was excluded due to an inability to maintain nominally sufficient driving control in the simulator. Demographic details per experimental condition are provided in Table 4.2.2. Across the randomly assigned groups, a large degree of demographic similarity was obtained (except for condition 4C where a lower proportion of females was represented).

Table 4.2.2. Overview of demographics per experimental condition (after one participant exclusion each from EXP3A and EXP4C).

Condition	Male	Female	Average age	Average driving frequency [#]	Average age first license
EXP1: continuous automation – no hands	8	5	23.46	3.85	18.97
EXP2: continuous automation – one hand	9	4	23.31	4.08	18.71
EXP3A: backup automation, eyes + situation, explicit	8	4	23.17	4.36	18.55
EXP3B: backup automation, eyes only, explicit	9	4	23.92	4.17	18.90
EXP4A: backup automation, eyes + situation, implicit	9	4	23.08	4.00	18.76
EXP4B: backup automation, eyes only, implicit	10	3	23.46	3.46	18.73
EXP4C: conventional driving, no automation	11	1	24.25	3.75	18.50

[#] 1 = every day, 2 = four to six days a week, 3 = one to three days a week, 4 = once a week to once a month, 5 = less than once a month, 6 = never.

Separate training exposure periods (about three minutes) were given for the driving simulation and the N-back task before simultaneous tasking was required in the experimental drive. After completion of the experimental drive, participants were presented with an on-screen response sheet that probed the participant's self-perception of the success and effort spent regarding aspects of safety, efficiency, and the N-back task, of the full driving trial, as well as (if applicable) satisfaction with the driving automation. The specific spatial layout and instructions of the subjective response sheet items are presented in Figure 4.2.3.

Subjective Response Sheet

Instructions:

Regarding the last driving trial session you just experienced,
Please TYPE IN a number for each of the items below

How (un)successful was the ... ?

(fail) 1 2 3 4 5 (success)

How little/much effort did you spend on ... ?

(low) 0 1 2 3 4 5 6 7 8 9 10 (high)

Safety:

Travel Time/Speed:

N-back Task Performance:

If you experienced automation in the last driving trial, ...

How (un)pleased were you with the automation?

(unhappy) 0 1 2 3 4 5 6 7 8 9 10 (happy)

Participant Number: p00

Condition Code: EXP00

SAVE

Figure 4.2.3. On-screen post-trial subjective questionnaire.

2.8. Driving route, timing and hazards

All experimental drives lasted about 2m45s. The route featured a straight road segment (0–40 s), a left curve (40–50 s), a straight road segment (50–70 s), a right curve (70–80 s), a straight road segment (80–120 s), a right curve (120–130 s), and a straight road segment (130–165 s). To test driver attentional engagement in the driving task, in each drive two surprise stationary obstacles were presented. The automation was programmed to drive through these objects as simulated detection errors. The objects had the form of a fallen tree and a stalled motorcycle (see Figure 4.2.4) presented at around 75 and 133 seconds, with response time-budgets of approximately 5 and 2 seconds, respectively. All of the aforementioned time descriptions are drawn from EXP1/2, in which the speed was computer-controlled (see Table 4.2.1); otherwise, speed variations affected the timing of route progress. The simulation had to be manually terminated because the driving automation implementation would not function beyond its scripted nominal trajectory time-series. A common data measurement cut-off point was established at 147.75 seconds (8865th frame at 60 Hz) for all seven conditions as this was the earliest point the simulation was manually terminated by the experimenter (i.e., participant 61 in EXP4A).



Figure 4.2.4. Stationary obstacles in the driving simulation appearing first as a fallen tree (left) after around 1 minute of driving and second as a stalled motorcycle (right) after around 2 minutes of driving.

2.9. Measures

Measures taken at the discrete hazards events. In EXP1/2, plots of steering and brake inputs were manually inspected for conventional driving activity (e.g., non-constant values) within the period between obstacle appearance and contact. In EXP3A/3B/4A/4B, the experimenter took subjective note of participant awareness of the obstacle within the same period, and the objective status of automation (on/off) and participant eye position (on/off screen) were recorded at the point of any contact.

Measures taken continuously across full trial. Visual distraction was measured as the percentage of time the DMS classified a state of visual distraction, registering a “1” for distracted after 3 consecutive seconds of looking off the driving simulation screen and a “0” for attentive after 4 consecutive seconds of looking on the driving simulation screen. NDRT performance was taken as a percentage of an aggregate final score at the end of a driving trial divided by the number of shown targets during that trial; one point was given for each correct response, and one point was subtracted for each incorrect or missed response. Automated driving status was measured as the percentage of time the vehicle was under automated control. Lateral performance was assessed as

road departures whenever the front left and/or right corner of the car was positioned above the grass area alongside the roadway. The car was 5.20 meters long and 2.03 meters wide, and the road was 6.4 meters wide with two lanes of 3.1 meters and two shoulders of 0.1 meters. Longitudinal route progress was calculated in meters traveled along the driving route. Perceptions of success (on a scale from 1 to 5) and effort (on a scale from 0 to 10) were each probed separately and in regards to the aspects of safety, travel efficiency (time/speed), and the N-back task performance at the end of each driving trial (Figure 4.2.3). Lastly, for all conditions containing a visible status display of automated control, participants were asked to rate their satisfaction on a scale from 0 to 10 (Figure 4.2.3).

2.10. Comparisons for each Research Question

Generally, the dependent measures could be captured and applied across the different research question comparisons (RQ1-5). However there were exceptions where certain measures would not make sense to apply, and some key measures had higher conceptual relevancy within a particular comparison than for others. For example, because the automation was active for the entirety of EXP1, measures of lateral and longitudinal control, as well as the proportion of time with activated driving automation, were not meaningful for this condition. Similarly, satisfaction with the automation could not be assessed for EXP4C (because this condition did not have any automation) or for EXP4A/4B (because participants were not told that this condition had any automation).

For RQ1, '*Are drivers susceptible to dangerous levels of distraction with SAE Level 2?*' the conditions EXP1 (Automation – no hands) and 4C (Conventional Driving – No Automation) were compared. The key objective measure here was the generation of a response to the hazardous obstacles, and the key subjective measures were perceived effort for time/speed efficiency of travel and perceived success with the N-back task. Supporting measures included the amount of objective visual distraction and N-back task performance.

For RQ2, '*Does placing a hand on the wheel improve driver supervision of automation?*' an improvement from EXP1 was sought by comparing EXP2 (Automation – one hand) with EXP1 (Automation – no hands). The key measures were thus the same as RQ1. Satisfaction with automation was also of interest regarding a potential detriment to end-user experience for having to keep one hand on the wheel.

For RQ3, '*Is adaptive backup a safe and acceptable alternative to continuous automated driving?*' a combination of all adaptive backup driving automation conditions (EXP3A/3B/4A/4B) was compared against a combination of all continuous automation conditions (EXP1/2), as well as against conventional driving (EXP4C). Key objective measures of interest included amounts of visual distraction and N-back task performance between EXP1/2 vs. EXP3A/3B/4A/4B, and the lateral performance measure of road departures between EXP3A/3B/4A/4B and EXP4C. Key subjective measures included perceptions of success/effort with the N-back task and perceived safety success/effort. Supporting measures included a measure of satisfaction with the automation, perceived success/effort spent on efficiency, and hazard collisions.

For RQ4, '*Can situated criteria safely reduce driver state monitoring from over-triggering?*' the set of situated adaptive automation conditions (EXP3A/4A) were compared against the set of human-centric adaptive automation conditions (EXP3B/4B). The key objective measure was the amount of

automated driving control and its consequential impact regarding efficient travel (longitudinal progress) in conjunction with safety (road departures). Key subjective measures were perceptions of success/effort for both safety and efficiency and satisfaction with the automation. Supporting objective measures included NDRT scores, amount of visual distraction, and hazard collisions, as well as subjective perceptions of success/effort on the NDRT.

For RQ5, '*Is the status of backup driving automation necessary to display to drivers?*' the set of implicit adaptive automation status conditions (EXP4A/4B) was compared to the set of explicit status conditions (EXP3A/3B). Visual distraction, NDRT performance, and proportion of automated control were key objective measures of over-reliance. Road departures and hazard collisions were key objective measures of safety. Trade-offs in perceptions of success/effort for safety vs. the NDRT were key subjective measures. Supporting measures included longitudinal progress performance and perceptions of success/effort for efficiency.

3. Results

Overall, our present experimental design included six objective dependent measures and seven subjective dependent measures and seven conditions as previously described. Data summaries of the measures across conditions are provided in Table 4.2.3, Table 4.2.4, Figure 4.2.5, and Figure 4.2.6. All inferential statistics are given in Table 4.2.5 for one-way ANOVA comparisons between EXP1 (no hands), EXP2 (one hand), and EXP4C (conventional driving); in Table 4.2.6 for t-test analyses to compare adaptive backup driving conditions as a set (EXP3A/3B/4A/4B) against continuous supervised automation conditions as a set (EXP1/2); and in Table 4.2.7 for two-way ANOVA comparisons between the different levels of backup design: assessment criteria (eyes-only vs. eyes-plus-situation) and interface display (explicit automation states vs. implicit automation status).

Table 4.2.3. Overview of responses made to hazard obstacles in the EXP1 and EXP2 conditions

n	Condition	Hazard Order, Content, Elapsed time	No response	Response: Steer only	Response: Brake only	Response: Steer & brake
13	EXP1: Automation – no hands	1 st , tree, 60s	10	2	0	1
13	EXP1: Automation – no hands	2 nd , motorcycle, 120s	2	7	1	3
13	EXP2: Automation – one hand	1 st , tree, 60s	2	11	0	0
13	EXP2: Automation – one hand	2 nd , motorcycle, 120s	2	10	0	1

Note: Non-response events were presently ambiguous in all experimental conditions containing some level of conventional control inputs due to inability to isolate steering and/or pedal inputs specifically intended for hazard avoidance (i.e., EXP3A/3B/4A/4B/4C).

Table 4.2.4. Overview of collisions and circumstances with hazard obstacles in the EXP3A, 3B, 4A, 4B, 4C conditions

n	Condition	Hazard Order, Content, Elapsed time	Collision	Conventional control (automation off)	Eyes Away (off- screen)	Not Tryin g to Avoid #
12	EXP3A: backup, eyes + situation, explicit	1 st , tree, 60s	11	2	1	2
12	EXP3A: backup, eyes + situation, explicit	2 nd , motorcycle, 120s	6	4	0	0
13	EXP3B: backup, eyes only, explicit	1 st , tree, 60s	4	0	4	0
13	EXP3B: backup, eyes only, explicit	2 nd , motorcycle, 120s	2	1	0	0
13	EXP4A: backup, eyes + situation, implicit	1 st , tree, 60s	9	1	2	1
13	EXP4A: backup, eyes + situation, implicit	2 nd , motorcycle, 120s	7	5	2	0
13	EXP4B: backup, eyes only, implicit	1 st , tree, 60s	6	2	3	0
13	EXP4B: backup, eyes only, implicit	2 nd , motorcycle, 120s	0	n.a.	n.a.	n.a.
12	EXP4C: conventional driving, no automation	1 st , tree, 60s	0	n.a.	n.a.	n.a.
12	EXP4C: conventional driving, no automation	2 nd , motorcycle, 120s	0	n.a.	n.a.	n.a.

* as per experimenter notes via subjective observation.

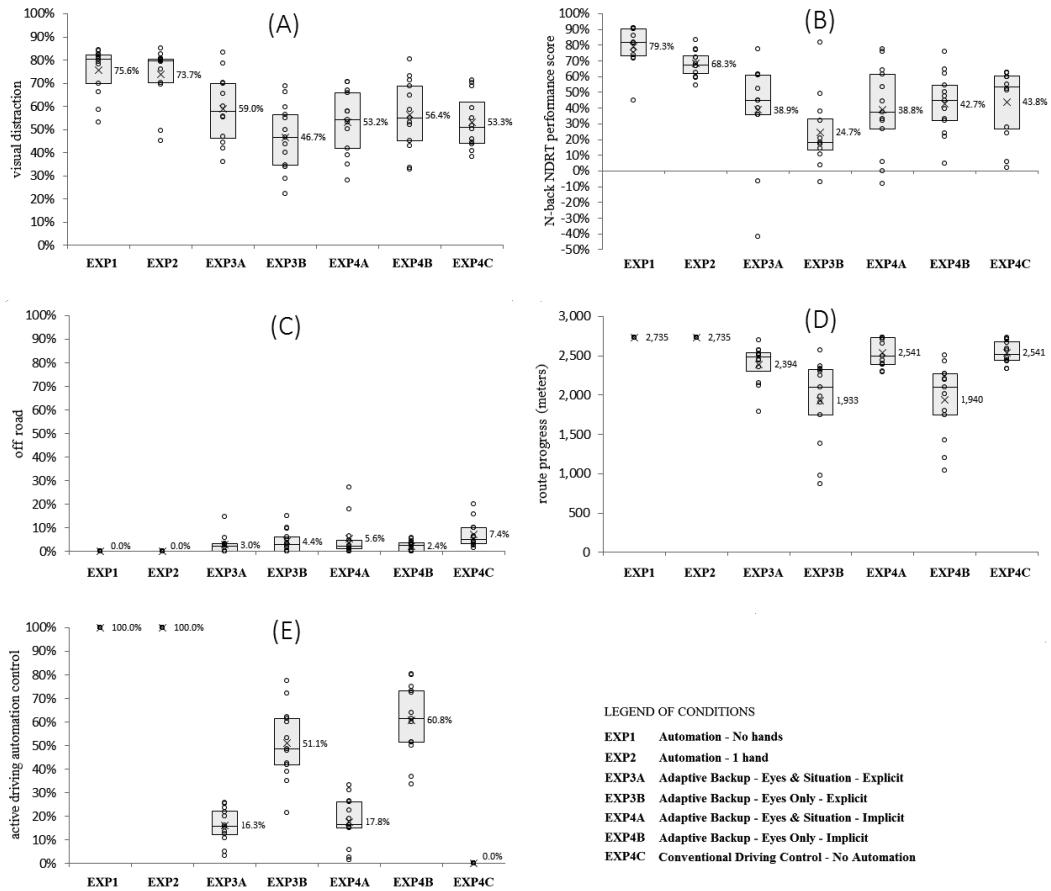


Figure 4.2.5. Objective results with means ('x'), medians ('—'), inner quartiles (‘◻’), and individual data points ('o') per condition for the measures of a) classified visual distraction, b) N-back NDRT performance, c) lateral performance, d) longitudinal performance, and e) amount of automated driving control. The numbers next to the boxplot represent the mean values.

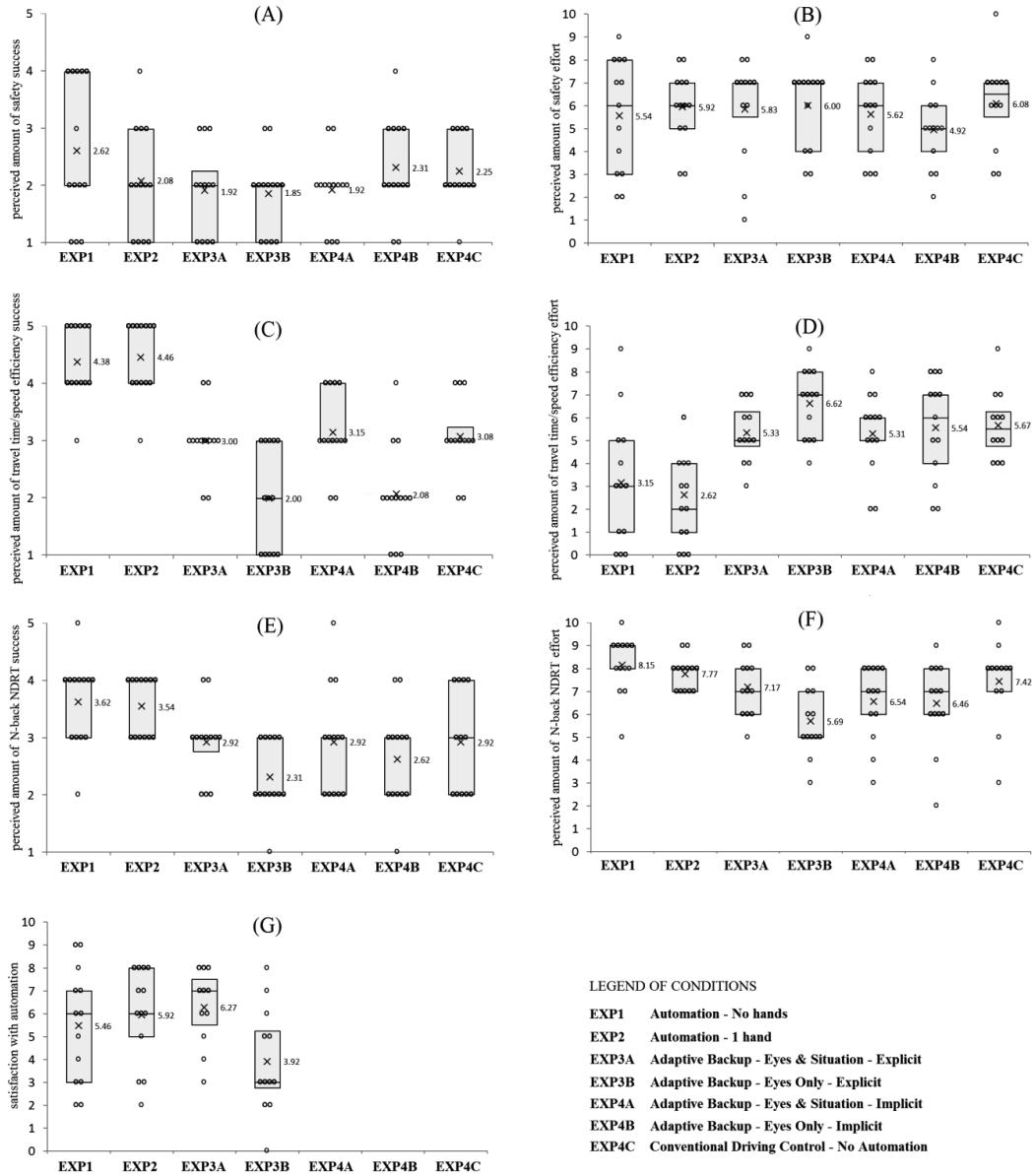


Figure 4.2.6. Subjective results with means ('x'), medians ('—'), inner quartiles ('◻◻'), and individual data points ('o') per condition for the measures a) safety success, b) safety effort, c) efficiency success, d) efficiency effort, e) N-back NDRT success, f) N-back NDRT effort, and g) satisfaction with automation. The numbers next to the boxplot represent the mean values.

LEGEND OF CONDITIONS

- EXP1 Automation - No hands
- EXP2 Automation - 1 hand
- EXP3A Adaptive Backup - Eyes & Situation - Explicit
- EXP3B Adaptive Backup - Eyes Only - Explicit
- EXP4A Adaptive Backup - Eyes & Situation - Implicit
- EXP4B Adaptive Backup - Eyes Only - Implicit
- EXP4C Conventional Driving Control - No Automation

Table 4.2.5: Statistics from one-way ANOVAs to compare EXP1 (no hands), EXP2 (one hand), and EXP4C (conventional driving). Directions (\downarrow/\uparrow) are indicated for differences found to be statistically significant after Bonferroni correction ($p = 0.05/3 = 0.017$). The NDT scores were lost for one EXP1 and one EXP2 participant.

Objective Measures										
Visual Distraction										
$F(2,35) = 14.122; p < 0.001; \eta^2 = 0.447$										
Conditions										
$EXP4C \downarrow$										
$EXP1 \uparrow$		$EXP2 \uparrow$		$EXP4C \downarrow$		$EXP1 \uparrow$				
M		75.6%		73.7%		73.7%				
SD		11.7%		12.5%		10.3%				
$p < 0.001$										
Safety - Perceived Success										
$F(2,35) = 1.002; p = 0.378; \eta^2 = 0.054$										
Conditions										
$EXP4C$										
$EXP1$		$EXP2$		$EXP1$		$EXP2$				
M		2.25		2.25		2.08				
SD		0.62		1.26		0.62				
$p = 0.630$										
Safety - Perceived Effort										
$F(2,35) = 0.233; p = 0.794; \eta^2 = 0.794$										
Conditions										
$EXP4C$										
$EXP1$		$EXP2$		$EXP1$		$EXP2$				
M		5.48		6.08		5.92				
$p = 0.788$										
Travel Time/Speed - Perceived Success										
$F(2,35) = 16.984; p < 0.001; \eta^2 = 0.493$										
Conditions										
$EXP4C$										
$EXP1 \downarrow$		$EXP2 \uparrow$		$EXP4C \downarrow$		$EXP1 \uparrow$				
M		3.08		4.38		3.08				
$p < 0.001$										
Travel Time/Speed - Perceived Effort										
$F(2,35) = 6.242; p = 0.005; \eta^2 = 0.263$										
Conditions										
$EXP4C$										
$EXP1$		$EXP2$		$EXP4C$		$EXP1 \uparrow$				
M		5.67		5.92		5.67				
$p = 0.025$										
N-back Task Performance - Perceived Success										
$F(2,35) = 3.293; p = 0.049; \eta^2 = 0.158$										
Conditions										
$EXP4C$										
$EXP1$		$EXP2$		$EXP4C \downarrow$		$EXP1 \uparrow$				
M		2.92		3.62		2.92				
$p = 0.002$										
N-back Task Performance - Perceived Effort										
$F(2,35) = 0.945; p = 0.398; \eta^2 = 0.051$										
Conditions										
$EXP4C$										
$EXP1$		$EXP2$		$EXP4C \uparrow$		$EXP1 \uparrow$				
M		7.42		8.15		7.42				
$p = 0.366$										
$p = 0.790$										
$p = 0.747$										

Table 4.2.6. Statistics from t-tests to compare EXP3A/3B/4A/4B (adaptive backup) and EXP1/2 (continuous automation). For lateral performance, the adaptive backup was instead compared to conventional driving (EXP4C). For automation satisfaction, additional comparisons are included between EXP1 (no hands) and EXP2 (one hand), and between EXP3A (eyes-plus-situation) and EXP3B (eyes-only). Directions (\downarrow/\uparrow) are indicated for differences found to be statistically significant.

Objective Measures					
Conditions	Visual Distraction		N-back Task Performance - Score		
	Adaptive Backup \downarrow EXP3A+3B+4A+4B	Continuous \uparrow EXP1+2	Conditions	Continuous \uparrow EXP1+2	Conditions
M	53.7%	74.7%	M	73.8%	M
SD	14.8%	11.3%	SD	12.3%	SD
Safety - Perceived Success					
t(37) = -1.413, $p = 0.166$, $d = 0.387$					
Conditions	Adaptive Backup EXP3A+3B+4A+4B	Continuous EXP1+2	Conditions	Continuous \uparrow EXP1+2	Conditions
M	2.00	2.35	M	2.55	M
SD	0.75	1.13	SD	0.92	SD
Safety - Perceived Effort					
t(75) = -0.301, $p = 0.764$, $d = 0.073$					
Conditions	Adaptive Backup EXP3A+3B+4A+4B	Continuous EXP1+2	Conditions	Continuous \downarrow EXP1+2	Conditions
M	5.59	5.73	M	5.71	M
SD	1.91	2.07	SD	1.78	SD
Automation Satisfaction (i)					
t(49) = -1.530, $p = 0.132$, $d = 0.429$					
Conditions	Adaptive Backup EXP3A+3B	Continuous EXP1+2	Conditions	Automation Satisfaction (ii)	Conditions
M	4.64	5.69	M	No Hands EXP1	M
SD	2.63	2.28	SD	5.46	SD
Automation Satisfaction (ii)					
t(24) = -0.509, $p = 0.615$, $d = 0.200$					
Conditions	Adaptive Backup EXP3A+3B	Continuous EXP1+2	Conditions	One Hand EXP2	Conditions
M	5.69	5.92	M	5.92	M
SD	2.28	2.50	SD	2.10	SD
Automation Satisfaction (iii)					
t(21) = 2.771, $p = 0.011$, $d = 1.157$					
Conditions	Eyes-Only EXP3B \downarrow	Eyes-plus-Situation EXP3A \uparrow	Conditions	Eyes-Only EXP3B \downarrow	Conditions
M	3.92	6.27	M	3.92	M
SD	2.31	1.68	SD	2.31	SD

Table 4.2.7. Statistics from two-way ANOVAs to compare EXP3A/4A (eyes-plus-situation) with EXP3B/4B (eyes-only) and to compare EXP3A/3B (explicit automation status) with EXP4A/4B (implicit automation status). Main effects are reported for each factor; no significant interaction effects were obtained. Directions (\downarrow/\uparrow) are indicated for differences found to be statistically significant. The NDRT score was lost for one EXP3B participant.

Objective Measures					
Visual Distraction			N-back Task Performance - Score		
$F(1,47) = 1.290, p = 0.262, \eta^2_p = 0.027$			$F(1,46) = 0.481, p = 0.492, \eta^2_p = 0.005$		
situated	eyes-only	implicit	situated	eyes-only	$F(1,46) = 1.434, p = 0.237, \eta^2_p = 0.030$
EXP	EXP3A+4A	EXP3B+4B	EXP	EXP3A+4A	EXP
M	56.0%	51.5%	M	38.9%	M
SD	14.4%	15.2%	SD	29.7%	SD
Lateral Performance - Road Departure			Longitudinal Performance - Route Progress		
$F(1,47) = 0.361, p = 0.551, \eta^2_p = 0.008$			$F(1,47) = 23.208, \mathbf{p} < 0.001, \eta^2_p = 0.331$		
situated	eyes-only	implicit	situated	eyes-only	$F(1,47) = 0.490, p = 0.487, \eta^2_p = 0.010$
EXP	EXP3A+4A	EXP3B+4B	EXP	EXP3A+4A	EXP
M	4.3%	3.4%	M	2470.6	M
SD	6.5%	3.8%	SD	225.5	SD
Subjective Measures					
Safety - Perceived Success			Travel Time/Speed - Perceived Success		
$F(1,47) = 0.561, p = 0.457, \eta^2_p = 0.012$			$F(1,47) = 22.556, \mathbf{p} < 0.001, \eta^2_p = 0.324$		
situated	eyes-only	implicit	situated	eyes-only	$F(1,47) = 0.278, p = 0.600, \eta^2_p = 0.006$
EXP	EXP3A+4A	EXP3B+4B	EXP	EXP3A+4A	EXP
M	1.92	2.08	M	3.08	M
SD	0.70	0.80	SD	0.64	SD
Safety - Perceived Effort			Travel Time/Speed - Perceived Effort		
$F(1,47) = 0.238, p = 0.628, \eta^2_p = 0.005$			$F(1,47) = 2.382, p = 0.129, \eta^2_p = 0.048$		
situated	eyes-only	implicit	situated	eyes-only	$F(1,47) = 1.265, p = 0.266, \eta^2_p = 0.026$
EXP	EXP3A+4A	EXP3B+4B	EXP	EXP3A+4A	EXP
M	5.72	5.46	M	5.32	M
SD	2.03	1.82	SD	1.55	SD
N-back Task Performance - Perceived Success					
$F(1,47) = 118.913, \mathbf{p} < 0.001, \eta^2_p = 0.717$			$F(1,47) = 4.235, \mathbf{p} = 0.045, \eta^2_p = 0.083$		
situated	eyes-only	implicit	situated	eyes-only	$F(1,47) = 0.497, p = 0.484, \eta^2_p = 0.010$
EXP	EXP3A+4B	EXP3B+4B	EXP	EXP3A+4B	EXP
M	17.1%	17.1%	M	2.92	M
SD	8.8%	8.8%	SD	0.81	SD
N-back Task Performance - Perceived Effort					
$F(1,47) = 3.029, \mathbf{p} = 0.058, \eta^2_p = 0.061$			$F(1,47) = 0.029, \mathbf{p} = 0.875, \eta^2_p = 0.001$		
situated	eyes-only	implicit	situated	eyes-only	$F(1,47) = 0.405, p = 0.600, \eta^2_p = 0.001$
EXP	EXP3A+4B	EXP3B+4B	EXP	EXP3A+4B	EXP
M	6.08	6.08	M	6.84	M
SD	1.49	1.53	SD	1.49	SD

3.1. RQ1

'Are drivers susceptible to dangerous levels of distraction with SAE Level 2?'

No collisions occurred in the conventional driving condition (EXP4C) for either the first or the second obstacle (Table 4.2.4). With automation without any hands on the wheel (EXP1), 10 out of 13 participants (77%) did not make any response to the first obstacle and 2 out of 13 participants (15%) made no response to the second obstacle (Table 4.2.3). Bonferroni-corrected alpha showed that perceived effort spent on travel time/speed was not significantly lower in EXP1 than in EXP4C (Table 4.2.5). Perceived success on the NDRT was not significantly higher in EXP1 than in EXP4C (Table 4.2.5). Objectively, participants exhibited significantly higher levels of visual distraction and improved NDRT scores in EXP1 than in EXP4C (Table 4.2.5).

3.2. RQ2

'Does placing a hand on the wheel improve driver supervision of automation?'

For the initial hazard, there were 10 non-responses in EXP1 (no hands) compared to 2 non-responses in EXP2 (one hand) (Table 4.2.3). However, non-responses to the second hazard were equally frequent (2 non-responses each) in the EXP1 and EXP2 conditions (Table 4.2.3). Perceived effort spent on travel time/speed and perceived success on the NDRT were not significantly different in EXP2 than in EXP1 (Table 4.2.5). Objective amounts of visual distraction and NDRT performance scores were also not found to differ significantly between EXP2 and EXP1 (Table 4.2.5). Satisfaction with the automation did not significantly differ between EXP2 and EXP1 (Table 4.2.6).

3.3. RQ3

'Is adaptive backup a safe and acceptable alternative to continuous automated driving?'

Visual distraction and NDRT performance scores were significantly lower in the set of adaptive backup conditions (EXP3A/3B/4A/4B) in comparison to the set of continuous supervised automated driving conditions (EXP1/2) (Table 4.2.6). Road departures were also significantly lower in EXP3A/3B/4A/4B compared to conventional driving (EXP4C) (Table 4.2.6). Participants in EXP3A/3B/4A/4B reported significantly lower effort and success with the NDRT compared to EXP1/2 (Table 4.2.6). Significant differences were not found between EXP3A/3B/4A/4B and EXP1/2 in regards to perceived safety effort, perceived safety success, or satisfaction with the automation (Table 4.2.6). For perceived travel time/speed, EXP3A/3B/4A/4B participants reported significantly higher effort and significantly lower success than EXP1/2 (Table 4.2.6).

Compared to the rate of non-response errors to hazards in EXP1/2 (16 of 52 possible, 31%) (Table 4.2.3), a lower rate was observed of participants not noticing or not trying to respond to the hazards in EXP3A/3B/4A/4B (3 out of 102 possible, 3%) (Table 4.2.4). Five hazard collisions occurred in the adaptive backup conditions with unobserved participant awareness. 37 other hazard collisions occurred in the adaptive backup conditions but with explainable causes rather than being attributable to complacency errors: 19 when the participant was observed to be actively trying to avoid the hazard (i.e., unsuccessful in regaining control from the automation), and 18 due to simulation artifacts of DMS-control malfunctions where there was an automatic system mismatch

between states of system-classified distraction and system-generated control authority (i.e., conventional driving allowed while being classified as distracted, or automated control retention while being classified as non-distracted).

3.4. RQ4

'Can situated criteria safely reduce driver state monitoring from over-triggering?'

In the eyes-plus-situation inattention assessment conditions (EXP3A/4A), the proportion of triggered automated backup control was significantly less than in eyes-only conditions (EXP3B/4B) (Table 4.2.7). Consequently, longitudinal progress was significantly greater in EXP3A/4A than in EXP3B/4B (Table 4.2.7). Objectively, no significant increase was observed for road departures in EXP3A/4A vs. EXP3B/4B (Table 4.2.7). Perceived success for travel time/speed was significantly higher with EXP3A/4A vs. EXP3B/4B without significant difference in terms of subjective effort for this aspect (Table 4.2.7). Participants in EXP3A reported significantly higher automation satisfaction than those in EXP3B (Table 4.2.6). Perceptions of effort for safety and success of safety did not significantly differ between EXP3A/4A and EXP 3B/4B (Table 4.2.7). Additionally, no significant differences were observed between EXP3A/4A vs. EXP 3B/4B in terms of the amount of visual distraction, NDRT performance scores, or perceived success/effort on the NDRT (Table 4.2.7). For hazard collisions where the participant was definitively observed as not trying to avoid the obstacle, all events transpired within the situated (EXP3A/4A) rather than the non-situated assessment conditions (EXP3B/4B) but were overall generally rare as an occurrence (i.e., 3 collisions out of 102 total exposures for EXP3A/4A and EXP3B/4B) (Table 4.2.4).

3.5. RQ5

'Is the status of backup driving automation necessary to display to drivers?'

In regards to objective measures suggestive of expected overreliance, visual distraction was not found to be significantly higher in the explicit adaptive status display conditions (EXP3A/3B) than in the implicit adaptive backup conditions (EXP4A/4B) (Table 4.2.7). NDRT performance scores, proportion of automated control, and longitudinal progress also were not found to be significantly higher with EXP3A/3B vs. EXP4A/4B (Table 4.2.7). Between EXP3A/3B and EXP4A/4B, no significant difference was found for the safety measure of road departures (Table 4.2.7) and no discernable patterns in evidently rare occurrences of hazard collisions where the participant was observed not to be not attempting to avoid the obstacle (i.e., 2 in an explicit condition: EXP3A, and 1 in an implicit condition: EXP4A) (Table 4.2.4). No significant differences were observed to evidence trade-offs between perceptions of success/effort for safety, travel time/speed efficiency, or the NDRT performance between EXP4A/4B and EXP3A/3B.

4. Discussion

4.1. Supervisory problems and design-level solutions

4.1.1. RQ1

'Are drivers susceptible to dangerous levels of distraction with SAE Level 2?'

Firstly, ahead of the investigated automation re-design aspects, a confirmation of problems was sought. The results from the conventional driving control condition (EXP4C) baselined a relative level of visual distraction from our N-back task (e.g., an average of around 53.3%) associated with poor lateral performance in our simulated setup (e.g., an average of around 7.4% time spent off-

road). From such a level of classified visual distraction, significant increases in both of our simulated SAE Level 2 driving automation conditions were evidenced: up to an average of 73.7% (EXP2 – one hand on the wheel) and 75.6% (EXP1 – no hands on the wheel). This increase in secondary task involvement (i.e., a significant improvement was also found in N-back scores between EXP4C and either EXP1 or EXP2), most likely accounts for our evidenced results of inadequate supervision, where 46.2% of our participants with continuous driving automation made no corrections to an unannounced hazardous automation failure. The subjective results for EXP1 compared to EXP4C, and for EXP2 compared to EXP4C suggest a prioritization of participants towards viewing the driving automation as convenience commodity (significant decrease in perceived travel time/speed effort with significant increase in perceived travel time/speed success; significant increase/trend in perceived secondary task success) rather than safety aid (mixed results regarding safety success/effort).

4.1.2. RQ (2)

'Does placing a hand on the wheel improve driver supervision of automation?'

RQ2 aimed to provide evidence for whether a requirement for hand placement might begin to address the above-identified problems in SAE Level 2 driving automation supervision. With a hand-on requirement, participants committed fewer non-response errors to first and second hazards (15%, 4 of 26) than those without hand placement stipulation (46%, 12 of 26). These results are in contrast with Naujoks et al. (2015), where significant performance differences were not found during critical events between hands-on and hands-off supervised driving automation conditions. However, Naujoks et al. (2015) reported a majority of drivers in their hands-off condition (120s interval allowed hands-free) had actually kept contact with the steering wheel. Notably, our EXP2 did not produce significant differences from EXP1 in terms of visual distraction, NDRT scores, or perceptions of success/effort, which suggests improved hazard awareness from hand-on requirements to be produced by mechanisms other than NDRT involvement or subjective value proposition (as seen between EXP1 and EXP4C). Physical hand-wheel contact might represent linked mind-body benefits that remind/prime a human operator towards conventional driving responsibility and steering activity. This explanation is consistent with our observation of steering to be the majority response (i.e., compared to braking) when responses were made.

4.1.3. RQ (3)

'Is adaptive backup a safe and acceptable alternative to continuous automated driving?'

Ironically, the requirement for humans to continuously supervise driving automation implies humans to have a greater capacity, across a larger operational envelope, which suggests humans are better fit for a majority driving role and should be periodically supported only as needed rather than replaced. RQ3 sought to replicate and extend benefits of the adaptive concept investigated in Cabrall et al. (2018) where humans drive conventionally with periodic automated backup support (e.g., when distracted). Our combined set of adaptive backup conditions (EXP3A/3B/4A/4B) evidenced significantly lower visual distraction and NDRT performance compared to the supervised continuous automation conditions (EXP1/2), and with significantly fewer road departures compared to conventional driving (EXP4C). Compared to EXP1/2, the subjective results suggest adaptive backup worked as intended by drawing participants *back* into the driving task (significantly lower perceptions of success with higher levels of effort in terms of travel time/speed efficiency) and *away* from the NDRT (significantly lower perceptions of success with lower levels of effort in NDRT performance). Additionally, satisfaction ratings with the simulated short exposure sessions of

driving automation were not found to be significantly lower (between-subjects) with EXP3A/3B/4A/4B compared to EXP1/2.

Some engineering issues were observed for eye-based re-claim of control from the automation during critical responses. Our unintentionally longer implemented requirements (on-road glance duration of 4 rather than 2 seconds) for establishing readiness of visual attention (see Methods section 2.4) is a likely explanation. Nevertheless, these results raise further design trade-off considerations (not investigated by the present paper). More/less strict ToC attention duration requirements might function in relation to the respective absence/presence of hazard(s). The circumstances of DMS ToC blocking of a hazard-aware human (although present in our simulation) would be conceptually rare in the real world as a combination of other rare events: driver distraction to a level requiring back-up, hazard presence, and false negative automation error. In contrast, continual human supervision of automated driving is expected to increase risks by a combination of increased likelihoods: operational duration of automated control, hazard exposure rate, vigilance decrement, and (illicit) uptake of NDRT with attentional capture. Overall then, in consideration of the discussed risks, adaptive automated backup appears to be a better conceptual driving automation choice than continually human supervised driving automation from a purely probabilistic perspective.

4.1.4. RQ (4)

'Can situated criteria safely reduce driver state monitoring from over-triggering?'

Our DMS was designed with an intended negative consequence for end-user inattention – where others on-market (e.g., Tesla Autopilot or GM Super Cruise) have used alarms or feature lockout, ours included an impedance to forward driving progress (i.e., slowing down). The human-centric eyes-only DMS conditions (EXP3B/4B) had significantly greater proportions of automated control and consequently more longitudinal impedance compared to the eyes-plus-situation DMS conditions (EXP4A/4B). Correspondently, participants expressed negative subjective experiences with significantly lower ratings on perceived travel time/speed success (EXP3B/4B) and automation satisfaction (EXP3B). Importantly, the conservative shift towards less automatic DMS triggers did not detract from safety: the perceived success of safety did not significantly decrease and lateral performance errors (i.e., road departures) did not significantly increase. In other words, the situated criteria functioned as hypothesized to reduce false alarms (e.g., avoid the 'cry-wolf' effect) while also not (dangerously) increasing misses with an overly strict criterion level.

4.1.5. RQ (5)

'Is the status of backup driving automation necessary to display to drivers?'

In conjunction with the potential rebranding and redesign of driving automation to serve as a punctuate safety rather than continual convenience commodity, there is an ethical manufacturer responsibility to attempt to deter potential end-user misuse. With aims to reduce risks of automation misuse such as from behavioral adaptation (see Martens & Jenssen, 2012) or mode confusion (Sarter & Woods, 1995), the lack of end-user awareness of backup automation existence/status in implicit backup conditions (EXP4A/4B) was not seen here to carry additional consequences (i.e., no significant detraction from positive measures nor significant addition to negative measures). Even though our short-duration simulated trials did not obtain direct positive evidence (e.g., significantly decreased visual distraction in EXP4A/4B), it is reasonable to expect (as motivated in the introduction section) that people might allow themselves to become distracted

more often, for longer periods of time, expecting that the vehicle can always successfully back them up. Promisingly, our results do suggest that the notification of backup driving automation and detected distraction events might not be necessary from a DMS and so can practically remain in the background.

4.2. Limitations

In terms of external validity, we wish to emphasize first that all exposure sessions in our experiment were targeted as fairly short (no more than a few minutes) distraction stress periods to evaluate different consequences of automation and DMS design concepts. Thus, multi-tasking challenges were assumed (as motivated in the introduction) and purposely induced by our procedures. Our present experimental results are thus only suggestive, with more naturalistic vigilance, with more rich/complex driving scene environments, and longer-term effects remaining to be investigated further elsewhere for replication, validation, and generalizability purposes.

It should also be noted that the present DMS-based adaptive backup driving automation concept was limited by some implementation problems in terms of computer network delays with the eye tracking as well as some rapid oscillations with the situated transitions of control. This means that if a participant was visually distracted, the automation recognized this sometimes several seconds later than intended. Thus, although participants who were visually distracted received backup support as intended (i.e., there was a strong correlation between the percentage of time that participants were visually distracted and the total time that the automation was 'on' in the eyes-only conditions, $r = 0.98$, $n = 26$ for the EXP3B and EXP4B conditions combined), participants in the explicit eyes-only EXP3B backup condition might not have been able to directly predict/understand when the backup automation turned on or off. Furthermore, indicative of unevenly distributed ToC, the average number of conventional driving to automation ToC events was actually higher in the situated assessment conditions (EXP3A/4A, $M = 51.7$) than in the eyes-only conditions (EXP3B/4B, $M = 7.2$) although being shorter lived with lower durations of applied distracted status (i.e., automation sustained as 'on').

Additionally, it should be cautioned that our driving simulation and NDRT are only artificial analogs (i.e., limited field of view, lack of realistic force feedback in steering, lack of vestibular motion feedback, etc.) of their real-life constituent representations – the simulated vehicle handling was anecdotally characterized as '*slippery*' and the N-back task might be more demanding/compelling than a real-life distraction such as a mobile phone chat message. Moreover, perceptions of risk (and hence risk-taking behaviors) are rarely commensurate between driving simulators and real-life roads.

The ecological validity of specific single off-road glance duration thresholds (e.g., around two seconds) is a controversial driver distraction topic, and research has suggested that further studies should be open to investigating more elaborated measures such as frequencies of repeated glances off-road (Liang et al., 2014), as well as in relation to durations of on-road glances (Kircher & Ahlstrom, 2009; Seppelt et al., 2017).

For all of the above reasons, the presently reported results should be interpreted in relative terms (ordinal comparisons between conditions) rather than absolute numeric values.

4.3. Conclusions

The present investigation demonstrated attentional susceptibilities in drivers tasked to supervise simulated full-time driving automation in the presence of a compelling NDRT. A requirement to maintain one hand on the wheel provided some benefit but still exhibited problematic rates of visual distraction and non-responses to hazards. Although the NDRT was tasked rather than voluntary, the depth of involvement was left free to each participant's own behavioral discretion. Consequently, we evidenced dangerous levels of distraction rather than uncompromised multi-tasking. Our results exhibited such automation over-reliance problems as possible for occurring in as short of time as a single minute.

Instead of focusing or leaving the problem as one of innate human limitations to attempt to correct, the present paper motivated an ecological approach for '*changing-the-machine-to-fit-the-man*' via redesign allocations of the same technology as adaptive backup. Overall, decreases in distraction (with the same NDRT) and consequential improvements to driving safety were evidenced from the adaptive backup conditions. Situated DMS criteria reduced unnecessary automatic assessments of distraction, and implicit automation status removed unnecessary risks for human misuse of automation (e.g., over-reliance).

Under controlled between-subject comparisons, we have shown preliminary feasibility without significantly reduced levels of acceptance compared to status-quo counterparts of supervised continuous automated driving and eyes-only distraction assessment (i.e., our new designs did not materialize evident conceptual deal-breakers). Our rather homogenized participant groupings were randomly assigned between conditions where very little presumably varied other than the manipulations of interest. However, further studies of within-subjects design, would strengthen a claim of achieved levels acceptance of our concepts and more targeted survey studies might best assess comparative acceptance/satisfaction at a broader level (e.g., intent to purchase).

4.4. Application

Our presently explored problems and solutions are germane to ongoing real-world automation design directions and decisions. Beyond the widely reported first AV fatality, there continue to be potential '*procrustean bed*' issues of '*fitting-the-man-to-the-machine*' which ironically can be obscured as user errors rather than system design drawbacks. For example, over two years later in June 2018, a safety driver of a Waymo autonomous vehicle caused an accident when he fell asleep and inadvertently activated a transition back to manual control (Griswold, 2018) — this incident was recorded with the CA DMV authority as a '*conventional mode*' rather than an '*autonomous mode*' accident.

A redesign concept for automated driving control, from continuous to backup, aims to support momentary irregular human errors rather than grossly replace all human driving authority (both responsible and reckless together) with technology that is still evolving rather than matured. If adaptive driving automation is pursued, further design considerations should involve how much information a DMS uses in its assessments and how much the driver needs to know about the system. Additional vehicle sensors (e.g., forward/side facing cameras and/or radar) can help DMS defer to driving scene contexts (i.e., of present collision and course deviation risks) prior to ascertaining driver states like distraction and consequential triggers for alerts and/or transitions of control. Such a layer is expected to provide a more situated human-like or graceful interaction style

through reduced false alarms and drops in perceived value by the driver. In terms of anticipated benefits of implicit driving automation applications, several safety systems already set a precedence of event-driven occurrence without explicit status indications or any knowledge required from the driver. With reduced risks of driver over-reliance, example background automotive safety functions include the priming of automated emergency braking, seatbelt tensioners, and electronic stability control systems.

Key points

- Complacency effects can occur with automated driving systems in as short as one minute of time. This may occur in spite of direct instruction requiring drivers to '*monitor and correct the automated driving for any dangers/errors*' and a recently experienced automated driving error.
- The provision to keep one hand on the wheel had a positive impact on generating a response to the first obstacle. However, non-responses to the second follow-on obstacle were equally present in both the no-hands and the one-hand-on-the-wheel automated driving conditions.
- All presently investigated backup driving automation conditions (whether with trigger criteria of eyes-only or eyes respective of driving scene/situations; and whether with hidden or overt transitions of control) were successful in reducing the amount of time spent off the road in comparison to a conventional driving control condition.
- An implicit backup automated driving system is expectedly harder to misuse than one with an explicit interface, and situated alerts have the potential to reduce negative impacts of false alarms such as reduced perceptions of self-success and overall satisfaction.

Acknowledgments

The research presented in this paper was conducted within the project HFAuto – Human Factors of Automated Driving (PITN-GA-2013-605817).

References

Bertoncello, M., & Wee, D. (2015). Ten ways autonomous driving could redefine the automotive world. McKinsey and Company.

Bhuiyan, J. (2018). Elon Musk says Tesla crashes shouldn't be front-page news because there are more human-driven fatalities: That's not an accurate comparison. *Recode*, May 17, 2018. <https://www.recode.net/2018/5/17/17362308/elon-musk-tesla-self-driving-autopilot-fatalities>

Bird, G. (2018). Jaguar Land Rover: Subtle support. *Vision Zero International*, January, 2018. pp. 32-36.

Butefisch, C.M., Davis, B.C., Wise, S.P., Sawaki, L., Kopylev, L., Classen, J. & Cohen, L.G. (2000). Mechanisms of use-dependent plasticity in the human motor cortex. *Proceedings of the National Academy of Sciences*, USA, 97, pp. 3661-3665.

Cabral, C.D.D., Happee, R., & De Winter, J.C.F. (2016a). From Mackworth's clock to the open road: A literature review on driver vigilance task operationalization. *Transportation Research part F: Traffic Psychology and Behaviour*, 40, pp. 169-189.

Cabral, C.D.D., Janssen, N., Goncalves, J., Morando, A., Sassman, M., & de Winter, J. (2016b). Eye-based driver state monitor of distraction, drowsiness, and cognitive load for transitions of control in automated driving. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*. Budapest, Hungary.

Cabral, C.D.D. (2017). Customizable N-Back GUI secondary task. [Computer script]. Zenodo. Retrieved from <http://doi.org/10.5281/zenodo.891531>

Cabral, C.D.D.* Janssen, N.* & De Winter, J.C.F. (2018). Adaptive automation: automatically (dis)engaging automation during visually distracted driving. *PeerJ Computer Science*, 4:e166; DOI 10.7717/peerj-cs.166

Carsten, O.M.J., Lai, F.C.H., Barnard, Y., Jamson, A.H., & Merat, N. (2012). Control task substitution in semiautomated driving: Does it matter what aspects are automated? *Human Factors*, 54(5), pp. 747-761

Classen, J., Liepert, J., Hallett, M. & Cohen, L. (1999). Plasticity of movement representation in the human motor cortex. *Electroencephalography and Clinical Neurophysiology*, Supplement, 51, pp. 162-173.

Classen, J., Liepert, A., Wise, S.P., Hallett, M. & Cohen, L.G. (1998). Rapid plasticity of human cortical movement representation induced by practice. *Journal of Neurophysiology*, 79, pp. 1117-1123.

Coppola, C. & Frank, B.J. (2018). Report: Uber driver was watching 'The Voice' moments before fatal Tempe crash. *Arizona Republic, AZ Central*, June 22, 2018. <https://eu.azcentral.com/story/news/local/tempe-breaking/2018/06/21/uber-self-driving-car-crash-tempe-police-elaine-herzberg/724344002/>

Cyganski, R., Fraedrich, E., & Lenz, B. (2014). Travel time valuation for automated driving: a use-case driven study. In: *Transportation Research Board 94th Annual Meeting* .

De Waard, D., Van der Hulst, M., Hoedemaeker, M., & Brookhuis, K.A. (1999). Driver behavior in an emergency situation in the automated highway system. *Transportation Human Factors*, 1(1), pp. 67-82.

Euro NCAP (2018). For the first time, Euro NCAP puts automated driving technology to the test. *Euro NCAP, Press Releases*, October 18, 2018. <https://www.euroncap.com/en/press-media/press-releases/testingautomation/>

Fung, B. (2017). The driver who died in a Tesla crash using Autopilot ignored at least 7 safety warnings. *The Switch, Washington Post*, June 20, 2017.

Glaser, Y.G., Llaneras, R.E., Glaser, D.S., & Green, C.A. (2016). Relationship Between Driver Eyes-Off-Road Interval and Hazard Detection Performance Under Automated Driving (No. 2016-01-1424). (SAE Technical Paper).

Godthelp, J., Milgram, P., & Blaauw, G.J. (1984). The development of a time-related measure to describe driving strategy. *Human Factors*, 26(3), pp. 257-268.

Goncalves, R.C., Quaresma, M., & Rodrigues, C.M.A (2017). Approaches for loss of vigilance in vehicle automation: a meta-analytical study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Austin, Texas, USA. October 9-13, 2017.

Griswold, A. (2018). Waymo's self-driving car crashed because its human driver fell asleep at the wheel. *Quartz*, October 2, 2018. <https://qz.com/1410928/waymos-self-driving-car-crashed-because-its-human-driver-fell-asleep/>

Hancock, P.A. (2013). In search of vigilance: The problem of iatrogenically created psychological phenomena. *American Psychologist*, 68, pp. 97-109.

Hancock, P.A. (2017). On the nature of vigilance. *Human Factors*, 59(1), pp. 35-43.

Hedlund, J., Simpson, H., & Mayhew, D. (2006). International conference on distracted driving: Summary of

proceedings and recommendations (2-5 October). Retrieved from <http://www.distracteddriving.ca/english/documents/ENGLISH-DDProceedingsandRecommendations.pdf>

Hickman, J. (2015). Driver distraction from electronic devices: Insights and implications. *Society of Automotive Engineers (SAE) webinar WB1140*, 06 May 2015. <https://www.sae.org/learn/content/wb1140/>

Hof, R. (2016). Toyota: 'Guardian Angel' cars will beat self-driving cars. *Forbes*, 08 April 2016. <https://www.forbes.com/sites/roberthof/2016/04/08/toyota-guardian-angel-cars-will-beat-self-driving-cars/#1fa141497f1>

Jamson, A.H., Merat, N., Carsten, O.M.J., & Lai, F.C.H. (2013). Behavioural changes in drivers experiencing highly Automated vehicle control in varying traffic conditions. *Transportation Research Part C*, 30, pp. 116-125

Kircher, K., & Ahlström, C. (2009). Issues related to the driver distraction detection algorithm AttenD. In: *First International Conference on Driver Distraction and Inattention*. Gothenburg, Sweden.

Kircher, K., & Ahlstrom, C. (2017). Minimum required attention: A human-centered approach to driver inattention. *Human Factors*, 59(3), pp. 471-484.

Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., & Ramsey, D.J. (2006). The impact of driver inattention on near-crash/crash risk: an analysis using the 100-Car Naturalistic Driving Study data. *Tech Rep. No. DOT HS 810 594*. National Highway Traffic Safety Administration, Washington, D.C.

Large, D.R., Banks, V.A., Burnett, G., Baverstock, S. & Skrypczuk, L. (2017). Exploring the behavior of distracted drivers during different levels of automation in driving. *Proceedings of the 5th International Conference on Driver Distraction and Inattention*. Paris, France.

Liang, Y., Lee, J.D., Horrey, W.J. (2014). A looming crisis: the distribution of off-road glance duration in moments leading up to crashes/near-crashes in naturalistic driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58, pp. 2102-2106.

Loukopoulos, L.D., Dismukes, R.K., & Barshi, I. (2009). *The Multitasking Myth: Handling Complexity in Real World Operations*. London: Routledge Ashgate Publishing.

Lu, Z., Coster, X., & de Winter, J.C.F. (2017). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics*, 60, pp. 293-304.

Mackworth, N.H. (1950). Researches on the measurement of human performance. Reprinted in H.W. Sinaiko (Ed.), *Selected Papers on Human Factors in the Design and Use of Control Systems* (1961). Dover Publications, New York, pp. 174-331.

Martens, M.H., & Janssen, G.F. (2012). Behavioural adaptation and acceptance. In A. Eskandarian (Ed.) *Handbook of Intelligent Vehicles*. London: Springer Verlag, pp. 117-138.

Mehler, B., Reimer, B., & Dusek, J.A. (2011). MIT AgeLab delayed digit recall task (n-back). Working Paper 2011-3B. MIT AgeLab, 28 June 2011.

Merat, N., Seppelt, B., Louw, T., Engstrom, J., Lee, J.D., Johansson, E., Creen, C.A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., & Keinath, A. (2019). The 'out-of-the-loop' concept in automated driving: proposed definition, measures and implications. *Cognition, Technology & Work*. <https://doi.org/10.1007/s10111-018-0525-8>

Michon, J.A. (1978). Dealing with Danger. Technical Report no. VK 79-01. Traffic Research Centre of the University of Groningen, The Netherlands. http://www.jamichon.nl/jam_writings/1979_dealing_with_danger.pdf

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*. New York: Plenum Press, pp. 485 - 520.

Naujoks, F., Purucker, C., Neukum, A., Wolter, S., & Steiger, R. (2015). Controllability of partially automated driving functions: Does it matter whether drivers are allowed to take their hands off the steering wheel? *Transportation Research Part F: Traffic Psychology and Behaviour*, 35, pp. 185-198.

NHTSA (2008). National motor vehicle crash causation survey. Report to Congress (Report No. DOT HS 811 059). Washington, DC: National Highway Traffic Safety Administration (NHTSA).

NHTSA (2013). Distraction detection and mitigation through driver feedback. United States National Highway Traffic Safety Administration (NHTSA), Report No. DOT HS 311 547A.

NHTSA (2018). Automated Driving Systems Voluntary Safety Self-Assessment (VSSA) Disclosure Index. <https://www.nhtsa.gov/automated-driving-systems/voluntary-safety-self-assessment>

Norman, D. (1990). The problem of automation: inappropriate feedback and interaction, not over automation. In: D.E. Broadbent, A.D. Baddeley, & J.T. Reason (Eds.) *Human Factors in Hazardous Situations*. Oxford, England: Clarendon Press.

NTSB (2017). Collision between a car operating with automated vehicle control systems and a tractor-semitrailer truck. National Transportation Safety Board, public meeting, NTSB/HAE 17-XX. Available online (last accessed Feb 12, 2019) at: <https://www.ntsb.gov/news/events/Documents/2017-HWY16FH018-BMG-abstract.pdf>

Parasuraman, R., & Manzey, D.H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), pp. 381-410.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39, pp. 230-253.

Parasuraman, R., & Wickens, C.D. (2008). Humans: Still vital after all these years of automation. *Human Factors*, 50(3), pp. 511-520.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, 38 (4), pp. 665-679.

Petermeijer, S.M., Abbink, D.A., & de Winter, J.C.F. (2015) Should drivers be operating within an automation-free bandwidth? Evaluating haptic steering support systems with different levels of authority. *Human Factors*, 57, pp. 5-20.

Rockwell, T.H. (1988). Spare visual capacity in driving – revisited: New empirical results of an old idea. In A.G. Gale, M.H. Freeman, C.M., Haslegrave, P. Smith, & S.P. Taylor (Eds.). *Vision in Vehicles II*. Amsterdam, Elsevier Science, pp. 317-324.

Rosen, C. (2008). The myth of multitasking. *The New Atlantis*, no. 20 (Spring 2008), pp. 105-110.

Ryu, J.B., Sihn, Y.K., Yu, S.B. (2013). Assessment of risky driving caused by the amount of time focused on an in-vehicle display system. *International Journal of Automotive Technology*, 14, pp. 259-264.

SAE (2018). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. SAE International, Surface Vehicle Recommended Practice, J3016_201806. https://www.sae.org/standards/content/j3016_201806/

Samuel, S., Fisher, D.L., 2015. Evaluation of the minimum forward roadway glance duration critical to latent hazard detection. *Transportation Research Board 94th Annual Meeting* (No. 15-1707).

Sarter, N.B., & Woods, D.D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors*, 37, pp. 5-19.

Scallen, S.F., Hancock, P.A., & Duley, J.A. (1995). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Applied Ergonomics*, 26 (6), pp. 397-403.

Seaman, S., Lee, J., Seppelt, B., Angell, L., Mehler, B., Reimer, B., 2017. It's all in the timing: using the Attention algorithm to assess texting in the NEST naturalistic driving database. *9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*.

Seppelt, B.D., & Lee, J.D. (2007). Making adaptive cruise control limits visible. *International Journal of Human Computer Studies*, 65(3), pp. 192-205.

Seppelt, B., Seaman, S., Lee, J., Angell, L.S., Mehler, B., & Reimer, B. (2017). Glass half-full: use of on-road glance metrics to differentiate crashes from near-crashes in the 100-car data. *Accident Analysis & Prevention*, 107, 48-62. doi: 10.1016/j.aap.2017.07.021

Sheridan, T.B., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of Human Factors and Ergonomics*, 1, pp. 89-129.

Simonite, T. (2017). Toyota is working on software that could prevent drivers from making deadly mistakes on the road. *MIT Technology Review, Business Insider*, 08 Mar 2017. <https://www.businessinsider.com/toyota-guardian-software-could-protect-drivers-from-deadly-mistakes-2017-3>

Stefan, K., Classen, J., Celnik, P., & Cohen, L.G. (2008). Concurrent action observation modulated practice induced motor memory formation. *European Journal of Neuroscience*, 27, pp. 730-738.

Strauch, B. (2018). Ironies of automation: Still unresolved after all these years. *IEEE Transactions on Human-Machine Systems*, 48(5), pp. 419-433.

TASS International (2019). Simulation of ADAS and active safety. *Home, Products & Services, Products, PreScan*. <https://tass.plm.automation.siemens.com/prescan>

Walton, D., & Thomas, J.A. (2005). Naturalistic observations of driver hand positions. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8, pp. 229-238.

Appendix A. Developed Driving Research Tools

A.1. Driver Monitor System: Interface Layout

Driving automation can be integrated with driver monitoring systems (DMS) to produce real-time adaptive and automatic transitions of control (ToC) (i.e., reducing human-machine interface requirements on manual button presses, gesture generation/recognition, vocal commands, etc., as well as relatively late cognitive processing requirements such as conscious human awareness of a need for a ToC). Many DMS can respond to different physiological driver measurements (heart, breath, sweat, brain, hands, head/face, body, etc.). Particular promise, however, is presumed from the measurement of eyes based on accounts of the importance of visual information demands in driving (cf. Sivak, 1996), as well as a continued reduction in form factors of cameras, which is favorable towards practical instrumentation considerations of decreasing intrusiveness and cost.

Eye-based DMS have been developed with different eye measure attribute states, but few combine several measurements and state classifications in a parallel hybrid manner, and fewer still towards direct integration aspects with adaptive driving automation ToCs. The DMS referenced in the present Chapter 4.2 has been shared as a Simulink model in an open-source repository at <http://doi.org/10.5281/zenodo.893325> and functions by processing eye-behavior data through three separate analysis streams to detect non-mutually exclusive sub-states of driver distraction, drowsiness, and/or cognitive overload (Figure 4.2.A.1). The classification parameters within each stream were derived from eye-tracking driving research but are purposefully grouped and arranged so as to facilitate easy visual programming for adjustments per different research needs. Furthermore, Figure 4.2.A.2 shows how feedback was directly incorporated in the model to visually overview in real-time how the system is arriving at its classifications of distraction, mental overload, and/or fatigue based on the currently defined parameters (i.e., values and time windows).

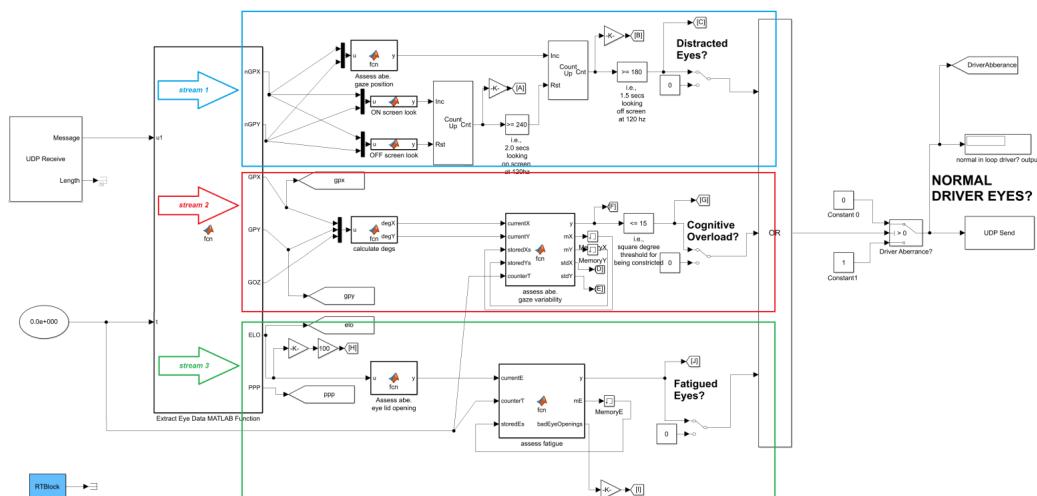


Figure 4.2.A.1. Beginning with UDP eye tracker inputs on the left, three separate yet parallel eye behavior analysis streams (from top to bottom: distraction, cognitive overload, fatigue) flow through the middle to the right where binary switch gates can be toggled to include consideration of the classified states towards an abstracted level of aberrance that can be transmitted via UDP as a single value for incorporation in an adaptive driving automation system's decision logic.

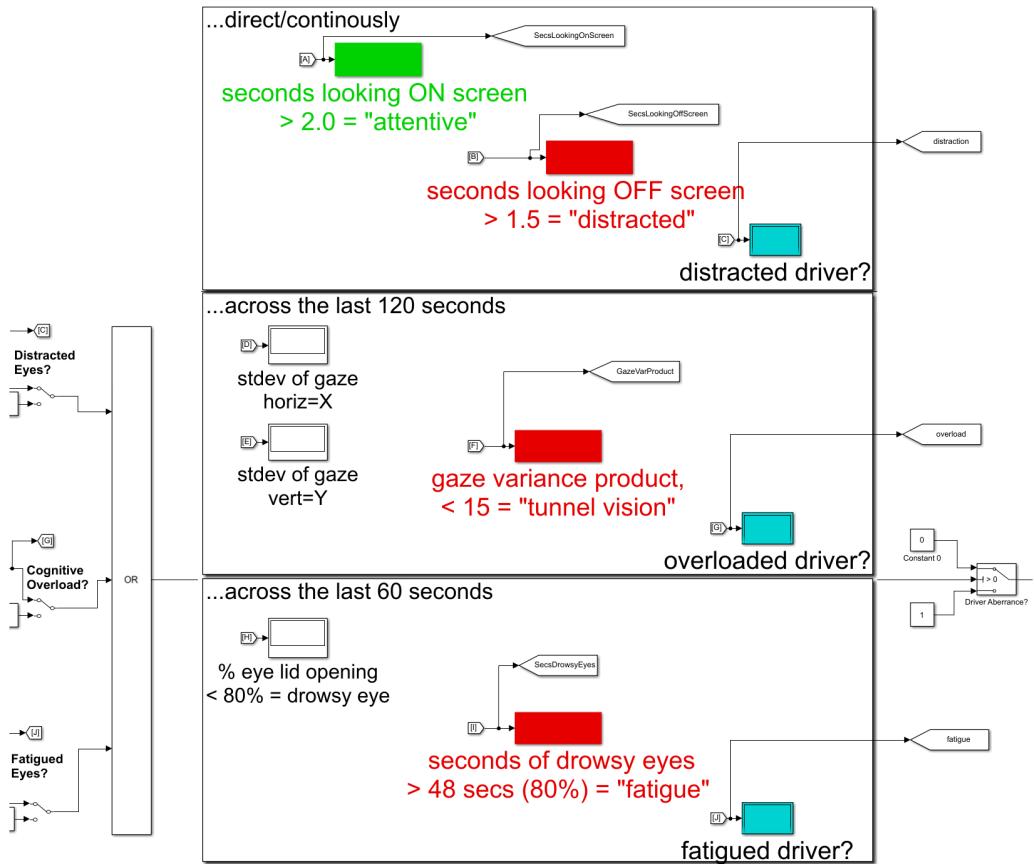


Figure 4.2.A.2. Additional direct visual feedback provided (e.g. for researcher/operator) for an overview of the currently defined classification parameters per the different state attributes (distracted, overloaded, fatigued) towards labelling the eyes of a driver as either aberrant or nominal.

A.2. Driving Automation Integration: Interface Layout

Packaged with the DMS model, is a Simulink model developed in conjunction with TASS International's PreScan physics-based driving simulation platform (see Figure 4.2.A.3). As an extension to the pre-existing automated driving control logic (top-middle) of lane center-ing at a set speed via linkages between vehicle states (middle-left) and vehicle dynamics (middle right), additional grouped block areas were self-implemented and organized to incorporate experimental control over adaptive lateral sensors for course conflict resolutions (top-left), adaptive longitudinal sensors for collision conflict resolutions (top-right), manual control (bottom-middle), data output (bottom-right) and functional allocation switches (middle-middle). For future studies, such a design facilitates direct manual experimenter control, in a pre-set or real-time fashion (via the circled binary switches) for what transitions of which driving control are (in)active and/or automatically driven by incoming eye classification data (bottom-left of middle-middle). The Simulink model is freely available within the online repository at <http://doi.org/10.5281/zenodo.893325>

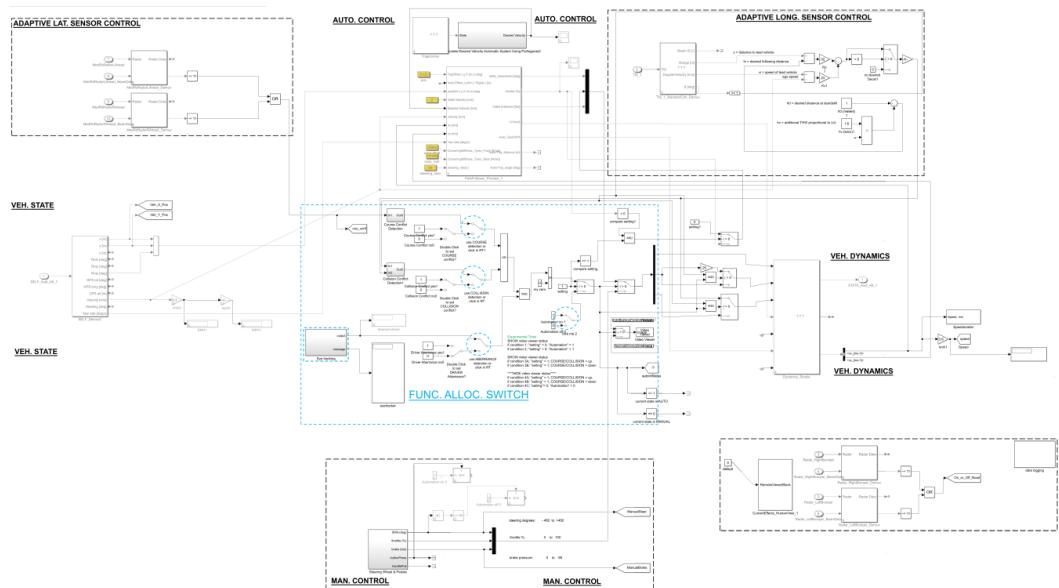
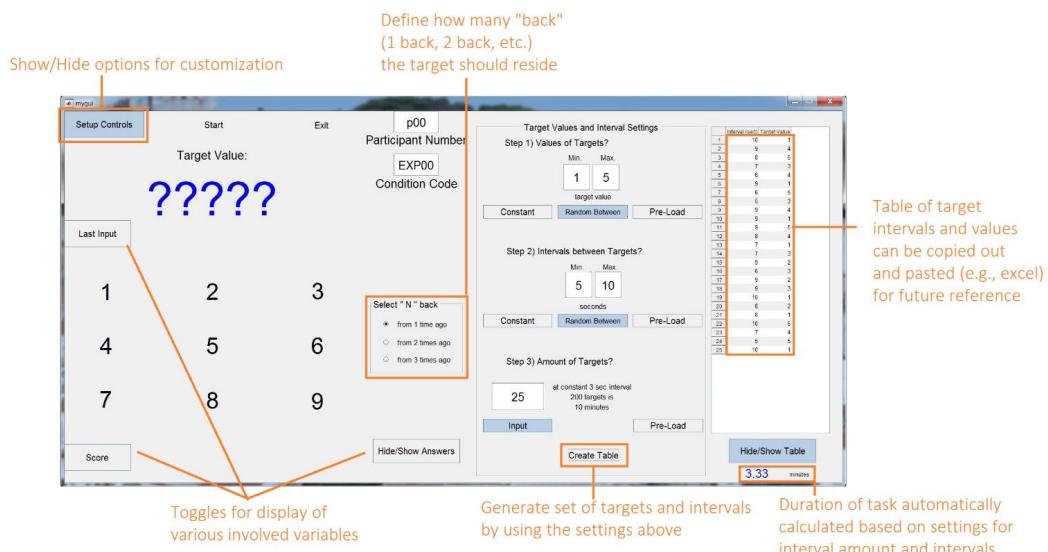


Figure 4.2.A.3. Extension of PreScan-Simulink model of driving automation platform for incorporation of automatic adaptive transition of control aspects towards facilitated experimenter/researcher control over the integration or isolation of different system components: automated lane centering and cruise control, manual steering and pedal inputs, lateral sensor triggers, longitudinal sensor triggers (with adaptive cruise control via acceleration suppression outputs), and/or eye-based classifications.

A.3. Visual N-Back GUI

A flexible programmable secondary task is provided as a modified version of N-Back. This GUI allows the experimenter to setup a visual manual N-back task that requires participants to key in responses when presented with a prompt “?????” asking them what was the target number seen 1 time, 2 times, or 3 times ago. Automatic scoring and auditory feedback is pre-programmed for correct and incorrect responses. The experimenter can customize the target values, intervals between targets, and amount of targets, and/or pre-load a provided set. Various display information items can be toggled on/off including: a running score, the last number the user responded with, the correct answer (whether from 1, 2, or 3 times ago), the table of target intervals and target values. From such customizable features, this N-Back secondary task can be adjusted in terms of difficulty/ease as needed (e.g., with more or less burden on memory). The Standalone executable and source code files are freely available within the online repository at <http://doi.org/10.5281/zenodo.891531>



5

PART 5: Discussion

Discussion chapter structure

This discussion chapter progresses the impact of the present thesis work first with an imperative and overview section. Subsequently, a summary of conclusions from each chapter is drawn out as a logical progression of individual studies and grouped part relations. The next section provides validating convergence of the present thesis study results with those from a few other recent theoretical, simulator, and on-road studies (that were all published in the years following completion of the present studies). A penultimate section provides a higher and lower level discussion regarding the bigger picture framework this thesis advances as well as where it specifically fits in for DMS applications. The discussion chapter concludes with a section pertaining to future research recommendations.

1. Thesis Imperatives and Impact Overview

Since the start of this thesis project in 2014, the consequences of overly simplistic conventions for assessing driver engagement have recently become all too real and deadly. In the wake of the first widely reported Tesla Autopilot fatality of Joshua Brown (May 17, 2016 in Florida), the U.S. National Transportation Safety Board (NTSB, 2017) issued new safety recommendations on September 12, 2017 for manufacturers to '*develop applications to more effectively sense the driver's level of engagement and alert the driver when engagement is lacking while automated vehicle control systems are in use*'. While crossing the street as a pedestrian, Elaine Herzberg was killed on March 18, 2018 in Tempe, Arizona, by an Uber '*self-driving*' test car equipped with a human safety driver who local police have reported was distracted by a streaming television program at the time (Plungis & Barry, 2018). Meanwhile, a recent National Safety Council public opinion poll (NSC, 2017) has found that drivers are actively disabling or otherwise defeating built-in safety features because they are either confusing, irritating, or susceptible to false alarms (Cichowski, 2017). Advanced driving assistance systems that rely on assessing driver attention through (periodic) steering wheel inputs have been subject to low-tech hacks from objects as common as an orange or a water bottle (Stumpf, 2018). Moreover, such defeat devices are even being openly sold as commercial products, e.g., the '*Autopilot Buddy* ®' from Dolder, Falco and Reese (2018).

What can be done about the deaths that are presently occurring on our roadways both from before and even still with driving automation? Improved human interactions with automatic driver monitor systems (DMS) should foster mutual calibrations of trust and ultimately benefit traffic safety through increased public adherence and appropriate use of its designed safety systems.

The aim of this thesis was '*to develop a system that is able to monitor the driver's vigilance*' and the approach taken was inspired by cognitive systems engineering (ecological perspectives). In-depth reviews of vigilance (Part 2) made it clear that situational knowledge would be crucial for understanding vigilance whether in general, for driving, or for monitoring driving automation. Furthermore, specific practical details with which to proceed to build a situated vigilance driver monitoring system (DMS) were found to be lacking. Thus, measurement studies (Part 3) were undertaken to better know the relevant details of driving scenes to which driver attention should appropriately relate. Amount of road curvature, traffic, and eye movement distances were identified as important and relatable factors.

Lastly, proof-of-concept integration studies in a driving simulator (Part 4) were deployed both with and without scene-tied assessment constraints. In the first version, control was taken away from human drivers and the vehicle slowed down whenever they looked away from the road too long; in the second version, looking away too long was tolerated by the system so long as no lateral or longitudinal conflicts were present. Benefits were obtained for both versions, but the situated DMS included a reduction of unnecessary alerts with improved primary and secondary task performance as well as enhanced participant acceptance ratings. In other words, the situated DMS respected natural human adaptive behavior (e.g., more secondary task involvement during less demanding driving) allowing them to better manage conflicting competition for attention.

Consequently, this thesis has succeeded in its approach to the stated objective and on-market DMS across levels of driving automation stand to be improved by incorporation of eyes and scenes taken together as a unified assessment. Beyond developing a single situated DMS, outputs of this thesis also were deliberately designed as several inroads towards extensibility for future research and development. When paired with AV technology, situated DMS will reduce unnecessary alerts to every instance of inadvertent supervisory attention over driving automation – instead focusing only on those that meaningfully matter such as when there are high visual demands presented by roadway curvature and/or increased traffic volumes. Thus, DMS can reach a level of social interaction intelligence that humans commonly expect when dealing with authority figures they more readily will comply with rather than reject or seek to undermine. When more people are able to use more driver monitoring and AV technology more appropriately more often, then road safety should reasonably be expected to increase.

2. Summary and Connection of Thesis Study Conclusions

Chapter 2.1: Driving vigilance task operationalization

Chap. 2.1 suggests the importance of vigilance tasking details (i.e., 18 are provided in Table 2.1.1) that are lacking for predicting/managing driving vigilance situations: specific consensus definitions of conventional driving signal(s), noise, and required response. Even with that same uncertainty, common visions for supervision of driving automation present greater risks of vigilance problems through an increase in overlap with other classic vigilance decrement features: temporal and spatial uncertainty (i.e., from manual de-skilling) of intermittent/rare signals (i.e., from growing reliability evolving automation) requiring time critical response (i.e., from take-over requests), within prolonged task durations (i.e., from enabling longer commutes/trips) and increased monotony (i.e., from computerized consistency in operation).

Vigilance is a pervasive topic. There were already around one thousand published reports on the topic by the mid 1980's (i.e., over 30 years ago). So a first place concern was to understand what has been known to cause vigilance decrements. Top-cited theory from across seven decades evidenced a list of around a dozen classic situational features (Table 2.1.1.) that were found to be highly contrived/constrained: as in consisting of specific signals (that must be few, temporally uncertain, short lasting, spatially uncertain, etc.), noise (that must be frequent and very similar to signals), and tasks (that must be long in duration, monotonous, and have required responses, etc.). Such artificial conditions evidenced as producing vigilance decrements is convergent with the titular 'iatrogenic' argument made by Hancock (2013) that 'locates the origin of the phenomenon and the onus for practical improvements ... with designers rather than apportioning blame for performance decrements to the operator ... (and) ... reinforces the recognition of ... the often unrecognized

external arbiter ... and the extrinsically imposed imperative to sustain attention'. In other words, if vigilance decrements were to be taken as a kind of disease it is arguably one that appears to be self-inflicted by design (i.e., by the specific operationalization of the vigilance task).

In prognosis of driving vigilance decrement issues, trying to map classic vigilance decrement situational features to the case of driving was determined to be a difficult, near impossible, endeavor. Too much uncertainty was present in reports of driving vigilance task operationalizations or else the signals, noise, and responses most commonly investigated were alongside of, rather than strictly belonging to driving percepts/actions (e.g., press a button upon hearing a 600 Hz but not a 500 Hz sinus tone). The difficulty in finding consensus operationalization of driving requirements at a specific/detailed level is probably best explainable upon reflecting that driving success (in the real-world) can be achieved in many different sufficient/satisficing rather than strictly optimized ways. However, even if specific definitions of driving signals, noise, and responses still remain unknown for operations involving supervising driving automation (just as with conventional driving), the overall supervisory task more closely approaches classic vigilance degradation situations by way of increased work constraints/pressures and reductive processes. What once was a complex/uncertain continuous task for the human driver, becomes a more simple/contrived intermittent task for the human supervisor of driving automation with: temporal and spatial uncertainty of intermittent/rare signals requiring time critical response, prolonged task durations, and increased monotony. Conclusions from Chapter 2.1 thus recommended caution and suggested (re)design opportunities against the status quo vision for deploying automated driving.

Chapter 2.2: Supervisory engagement with driving automation

Chap. 2.2 shows that the most common solution areas to the problem of keeping attention while supervising automation include those focused on internal cognitive states, followed by those with a broader situational (task/ecological) perspective.

Outside of recent developments in driving automation, increases in automation have been changing human roles/responsibilities from lower-level operators to higher-level supervisors in variety of domains for an extended period of time already. Consequently, there is a substantial body of human-automation interaction literature with concerns and suggested solutions for keeping up engagement/attention of human supervisors of automation. Chap 2.2. developed a categorization scheme of six themes to group the solutions into recognizable areas such that frequencies and trends analysis could be supported and applied. The first three themes describe supervisory control avoidance either in a hard sense or different versions of a soft stance: objective or subjective reductions in the supervisory control task. The latter three themes describe solutions under familiar learning theory paradigms in chronological order: behaviourism, cognitivism, and ecological constructivism. Results from Chapter 2.2 showed that independent raters were able to reliably apply the themes to categorize recommendations from influential human-automation interaction research. Cognitive followed by ecological themed solutions appeared to be the most commonly proposed in influential human-automation interaction literature conclusions. Additionally, less common but still evident areas suggested either avoiding the supervision task outright or ways to reduce it

Part 2: Driver Vigilance Review – take-away

Taken together, the studies of Part 2 emphasize the importance of cognitive and situational themed conclusions for managing vigilance issues in general, but a lacking of available practical details (i.e.,

what driving scene features and driver eye measurements) with which one might proceed to build a situated DMS. Thus, applied driver eye and driving scene measurement studies were conducted in Part 3.

Chapter 3.1: Crowdsourced driving scene content categorization

Chap. 3.1 produced a broad yet efficient driving scene content categorization scheme and confirmed relatively high levels of accuracy and reliability in crowdsourced annotations using that scheme. Thus, measurement of driving scene aspects was nailed down in a concrete and viable manner.

After the review work of Part 2, we faced the question of how driving scenes could be measured/described with a balance of comprehensive coverage and efficient annotation. Traffic safety literature suggests that driving scene situational features of general interest might fall under three categories of road users (and their locations), their behavior, and road/infrastructure details. For ease of annotation, items were strictly operationalized as only binary values for presence/absence (check boxes) and ordered in a probabilistically prioritized manner (more likely - first, less likely - later). Consequently, a single driving scene annotation of around 36 scene features took on average 37 seconds to complete. Several relatively easy/unambiguous driving scenes were pre-categorized and used as explicit training material as well as mixed in (in places unknown to crowdsourced annotators) as implicit screening devices to remove indiscriminate/incorrect responders. A robust (valid and reliable) driving scene library was thus able to be constructed consisting of about 38,298 seconds of dash-cam driving footage with their contents annotated by around 200 crowdworkers from 46 countries in about 1 ½ days' time.

Chapter 3.2: Prediction of workload, attention and eyes from driving scene contents

Chap. 3.2 determined specific driving scene features (i.e., road curvature and traffic) to be of importance to perceived driving effort ratings and associated eye movements (i.e., saccade amplitude).

Because some driving scenes are easier/harder than others, a situated DMS should be able to know how much attention to expect from a driver's eyes relative to such present demands to be more conservative/judicious in its vigilance assessments and involvement. So the measurement study of Chap 3.2. sought to determine what driving scene features would be associated with what eye measures (and in accordance with a range of perceived effort that drives those eye behaviors). The high volume of annotated scene segments in Chapter 3.1 (~12,862 scenes from around 50 different driving videos) enabled a selection of stimulus material that contained a sufficient degree of resolution to perform predictive regression analyses in Chapter 3.2 (i.e., continuous scaled independent variables to match continuous scaled dependent variable constructs). Specifically, 60 video clips were selected to represent a range of low/high driving scene demands with different scene features.

The most powerful relations were found for effort ratings as predicted from road curvature and traffic; saccade amplitude as predicted by effort ratings; and saccade amplitude as predicted from road curvature and traffic. More/less road angle curvature (and more/less traffic) was associated with more/less effort and more/less saccade amplitudes. Thus, the lower level eye movement measurements showed stronger (more reliable) relations with perceived effort and visible scene contents (lateral/longitudinal conflicts) than the higher level representation (and eye

measurement) aspects of information uptake (fixation duration) and increased cognitive processing (pupil size).

Chapter 3.3: On-road out-of-the-loop drivenger eyes

Chap. 3.3 measured both on-road eye movements and driving scene aspects. '*Out-of-the-loop*' eyes generally exhibited greater off-center movement distances across entire trips. However, the off-center distances of '*in-the-loop*' eyes were observed to periodically rise and fall with respectively low and high driving scene demands (as operationalized by steering angle, traffic count, and speed).

Within an on-road study environment, Chap 3.3. investigated a different characterization of eye-scene relations than was able to be determined in the laboratory environment of Chap 3.2 (where scene demands could be more precisely measured and safely manipulated). An eye-measurement difference was captured between the variant role/responsibilities of on-road drivers (who, by definition, are in control of the driving) vs. on-road passengers (who, by definition, are not in control of the driving). Benefits of this innovative approach included increased safety and naturalism when compared to more common research methods of imposing artificial distraction tasks to ensure the driving participant becomes '*out-of-the-loop*' (i.e., for the sake of making measurements at such points). An additional benefit was that paired participants served as comparative controls for one another in terms of being in the same vehicle as it moved between varying driving scene demands (traffic, weather, road-infrastructures, etc.).

Both driver and passenger eyes moved substantial distances on/off road center and around/across the driving scene. Across a driving trip as a whole, passenger eye eccentricity typically exceeded driver eccentricity (by about 25%). However, when driving scene demands were higher (increases in steering angles, traffic, and/or speed) discrimination performance weakened because driver eye eccentricity adaptively increased to meet those increased demands whereas passenger eye eccentricity was more free to vary in such situations. Driver eye eccentricity also rose during low demand situations where they became (like passengers) more free to vary. In conclusion, recommendations were made to discard DMS alerts to increased driver eye movements that reflect natural/safe adaption to relative extremities of high/low (visual) demands.

Part 3: Driving scenes and driver eyes – take-away

Taken together, the studies of Part 3 emphasize the viability of measuring relations between driver eyes and driving scenes at a behavioral level. An applicable situated DMS conclusion was that specific measureable (visible) scene demand features of road curvature and traffic count could reliably be represented in low-level pre-cognitive eye movement measurements. Next, the studies of Part 4 executed driving simulator proof-of-concept design validations of various integrations of real-time vigilance DMS and driving automation.

Chapter 4.1: Directionality of eye-based transitions of driving control

Chap. 4.1 implemented a driving simulator proof-of-concept real-time DMS and driving automation integration (i.e., where the automation backs up a driver that looks away too long) that showed safety and acceptance improvements over an emulated concept of present-day on-market functional allocations of automated driving (i.e., where the automation de-activates itself upon detecting distraction).

As discussed in the literature reviews/surveys of Chap 2.1 and Chap 2.2, there are evident human factors concerns with a level of driving automation that requires human supervision (as backup/fall-back). Nevertheless, such systems have been released on public roads and the most popular conceptual instantiations are trending towards attempts to manage supervisory driver inattention issues by automatically disengaging themselves if driver engagement assessments are negative ('*forced-manual control*'). The experiment of Chap 4.1 investigated safety implications of such an integration implementation against a role reversed concept ('*adaptive-backup control*') where the driving automation instead backs up the human driver upon inattention assessments.

Peak absolute lateral error was higher with the forced-manual control condition compared to the adaptive-backup control condition. The adaptive-backup control condition showed lower self-reported workload ratings and yielded higher acceptance ratings than the forced-manual control condition. Thus, driving performance and experiences were improved by reversing the directionality of the adaptive transition of control, i.e., keeping the majority/continuity of driving control with human drivers and backing them up with automated driving control when they become distracted.

Chapter 4.2: Situated/Implicit backup driving control

Chap. 4.2 extended the successful proof-of-concept from Chap 4.1 within another driving simulator study. Inattention problems with supervising driving automation were evidenced (but also reduced from a condition requiring one hand be kept on the wheel). Situated and implicit DMS integration designs of adaptive-backup control showed user interaction and performance improvements.

First, Chap 4.2 confirmed previously assumed inattention issues with supervision of driving automation by showing higher incidences of non-response errors to unexpected road hazards (as compared to a condition with full-time manual control and several versions of adaptive-backup control). Requirements with a low-level physical tie-in (i.e., keep one hand on the wheel) significantly improved hazard response generation.

Second, Chap 4.2. examined different versions of the successful implementation from Chap. 4.1 in order to examine further design improvements aimed at potential drawback issues of over-alerting and driver over-reliance (misuse) as might be problematic for adaptive-backup control. Situated DMS backup control (off-road looking with present lateral and/or longitudinal conflicts) generated higher perceptions of success, while reducing over-alerting without impacting safety from its lowered amount of involvement compared to non-situated DMS (off-road looking only). The implicit design where adaptive backup control status indication was removed/hidden did not produce any disadvantages.

Part 4: Adaptive driving automation – take-away

Taken together, the studies of Part 4 emphasize problems with presently released driving automation designs where humans supervise without continuous physical activity involvement requirements. Most importantly, the Part 4 studies confirm viability of real-time eye-based DMS integration with driving automation towards practical user experience and safety advantages not only when deployed in an adaptive-backup directionality for transition of control, but also as from a situated version of DMS specifically.

3. Recently Convergent Research

Since the completion of the design and conduct phases of the enclosed thesis work, an appreciable volume of publications within the last couple of years (2017 and 2018) have meanwhile emerged that appear to be in agreement with the present thesis topic, theoretical aims, methods, and/or results. A few of such are taken as examples for discussion in this section pertaining to validating convergence of my thesis research with the research of others.

3.1 Predicted problems and solution directions

Automation-related concerns have been a main-stay in human-machine research for decades before and are predicted to continue. Strauch (2018) indicated that ironies such as introduced from Bainbridge (1983) are remaining rather than resolved even while automation is proliferating well beyond the professional operator settings of previous generations (e.g., now ubiquitous in unregulated public arenas as with smartphones and automobiles rather than nuclear power processing plants). In the driving automation domain, he cited driving simulator research indicating delayed reaction times and reduced attention of vehicle operators in highly automated driving compared to manual control. From the present thesis, Chapter 2.1 concluded with design features of driving automation systems (e.g., prolonged periods with low-frequency signals, signals that are similar to noise, lack of feedback on performance, etc.) that suggest an increased likelihood of classic vigilance problems.

In terms of recommendations, Strauch (2018) suggested further research that examines how drivers can retain skills enabling them to effectively recognize and respond to critical situations. Strauch suggested systems that retain the features of automation (safety, reliability, accuracy, economy) while at the same time optimizing human drivers' vigilance and retention of manual and cognitive operating skills. Presently, Chapters 4.1 and 4.2 have found benefits from using driving automation as a backup to humans (e.g., a reverse of the typically promoted proportions of human/automation driving control), an approach that may maximize the safety strengths of automation while retaining operator manual driving skills. Furthermore, Strauch endorsed a solution from Bainbridge as 'worthwhile', where operators are given opportunities to practice manual control during actual system operations, and if not possible, by providing similar experiences in system simulators (e.g., consistent with theme #2 of the present Chapter 2.2). He concluded with an emphasis on provisions of training to meet the additional technology (e.g., consistent with theme #4 of the present Chapter 2.2), which he argued is important given the problem of increases in automation in non-professional domains (e.g., automotive).

3.2 Performance and attention measures in driving automation simulator studies

Greenlee et al. (2018) tasked participants to press a button on the steering wheel upon detection of low probability (5%) roadway hazards (i.e., vehicles encroaching upon the lane of travel) without feedback while supervising an automated driving vehicle for trips of about 40 minutes under foggy conditions. Across subsequent 10 minutes periods of watch, correct detections declined in number (i.e., more than 30%) with a significant drop evidenced onward from the 10-20 minute time-on-task period. Likewise, reaction times were found to significantly increase after the first period of watch. Post-drive workload self-ratings descriptively showed above scale mid-point average ratings for mental demand, temporal demand, effort, and frustration and a significantly higher than mid-level

global demand. Pre- and post-drive self-reported stress ratings indicated significantly decreased engagement and significantly increased distress. Comparably, Stapel et al. (2019) found on-road automated driving to reduce perceived workload, but monitoring duties therein to increase cognitive workload when compared to conventional/manual driving.

The vigilance task operationalization of Greenlee et al. (2018) appears highly in overlap with the composite multi-decade vigilance decrement set of features identified in the present chapter 2.1 (e.g., rare and difficult to perceive signals that are similar to frequent noise in a prolonged monotonous task without feedback, etc.). Recommended augmentative strategies to researchers and developers of vehicle automation strategies included breaks from the monitoring duties and use of physiological monitoring to continually assess and adaptively respond to measured driver vigilance. Both of these recommendations are consistent with theme #2 of the present chapter 2.2 and are supported by the eye tracking measures of Chapter 3.2 and 3.3 as well as the incorporated DMS development and application in chapters 4.1 and 4.2. In Chapter 4.2, human vigilance inadequacies in supervising driving automation were found in time periods as short as 1 minute (and even upon following a recent automation failure) when exacerbated by a compelling secondary task.

3.3 Real-world driver SA and behavior issues with released on-road driving automation

Endsley (2017) conducted a 6 month longitudinal study of personal naturalistic experiences with the driving autonomy features of her own Tesla Model S. While SA (as measured from real-time knowledge probes) was not found to be significantly higher or lower on average than a control period, her observation was that it was still problematically consequential for increased accident risk in being more variable and hence susceptible to being gone when it might be needed. Endsley (2017) perceived her reaction times to be slower: '*I was surprisingly slow to react ... it took extra seconds to realize that the automation was not going to handle the situation*'. With a secondary task, significantly increased visual distraction and significant non-responses to automation failure events were found in the driving simulator study of the present chapter 4.2 that emulated driving automation conditions similar to the use of Tesla Autopilot features (i.e., simultaneous automated driving lateral and longitudinal control with stipulations for visual/mental involvement and some variation in hands-on requirements).

Endsley (2017) also experienced false alarm problems with warnings that occurred '*frequently in error, causing significant frustration*' and stated dissatisfaction with a lack of face-validity in vigilance assessment: '*having one's hands on the wheel, is not the same as having one's mind on the road*'. Present thesis results contributed to the reduction of potential false alarms (Chapter 3.3) and their consequences (Chapter 4.2) in assessment of driver engagement and constructs of driver SA through eye-based measures (Chapters 3.2, 3.3, 4.1 and 4.2) rather than steering wheel input sensors. However, it should be noted that the experiment of the present chapter 4.2 also contributed counter-evidence that one-hand on the wheel in fact increased the likelihood of generating a response to driving automation failures compared to no-hands on the wheel. Endsley (2017) summarily states that the autopilot mode of Tesla '*is likely to provide a good backup*' and is consistent with the designs investigated in the present chapters 4.1 and 4.2. Lastly, Endsley (2017) advocated for increased driver training to address the new responsibilities with driving automation (consistent with the present chapter 2.2 theme #4) and also proposed many improvements of her

identified system design/interface issues towards supporting drivers' mental models and understanding of the automation (theme #5).

Banks et al. (2018) analyzed video observations they collected during an on-road study using a Tesla Model S being operated in Autopilot mode (i.e., 12 participants, approximately 40-minute driving trips each). Only one participant was observed to remain '*hands-on*' throughout their use of the Autopilot features (note: Tesla documentation states '*Autosteer is a hands-on feature. You must keep your hands on the steering wheel at all times.*'). Multiple warnings from the remaining drivers resulted in periods in excess of 75 seconds of '*hands free*' driving, which the authors lament as a substantial time period that could enable non-driving related secondary tasks to be taken up and might have '*disastrous consequences*' if at the same point in time an operational design domain (ODD) breach were to occur (e.g., an automation failure/error). Banks et al. (2018) cited (NHTSA, 2017) where the infamous Tesla Autopilot fatality (Joshua Brown) was attributed to a prolonged period of distracted driving. With both, one and no hands, on the steering wheel (and a compelling secondary task), the present chapter 4.2 found significantly increased instances of visual distraction when participants were supposed to be monitoring simulated SAE level 2 driving automation as compared to manual driving (also with the same secondary task). Furthermore, the emulated SAE level 2 conditions of Chapter 4.2 produced non-responses to simulated driving automation errors (i.e., driving into a fallen tree and through a motorcyclist) while participants were just so visually distracted as warned above by Banks et al. (2018).

Banks et al. (2018) observed substantial issues with mode confusion, visual human machine interface status displays, and false alarms. The present chapter 4.2 included system integration designs for combining human and driving automation to reduce the first two via implicit backup driving automation and the last via the incorporation of more situated automatic DMS assessments. Banks et al. (2018) concluded that either the human driver should remain in control of at least one of the control aspects (longitudinal and/or lateral) or they are removed entirely from the control-feedback loop thus skipping the middle SAE levels of driving automation involving supervisory driver control. Such a recommendation is consistent with the present chapter 2.2 theme #1.

3.4. Convergence summary

In recently published research of the last couple of years (2017 and 2018) problems have been identified for driver engagement (attention, vigilance, SA, etc.) and its assessment across various levels of driving automation. Suggested detailed understandings of the underlying issues are consistent with those identified in the literature review of the present Chapter 2.1 and offered solutions are convergent with themes discussed in Chapter 2.2. Additionally, continuous eye-based measures are being proposed and pursued both from information processing frameworks (e.g., internal focus on interpreting cognitive states of individual drivers) as well as in relation to broader external contexts. The present Chapters 3.2, 3.3, 4.1, and 4.2 all provide viable inroads to making use of eye-tracking data, while relations to driving scene situations were more directly considered in Chapters 3.2, 3.3, and 4.2. Furthermore, Chapter 4.2 contributed an integration design and implementation platform that could be useful for further researchers of similar interests which allows for preliminary/prototypical investigations of adaptive driving automation by means of a easily re-configurable DMS and driving control (i.e., via GUI toggle switches and/or numeric entry fields).

4. Thesis Research and Development Implications

4.1 Ecological theory framework

The research within the present thesis substantiates triadic theoretical paradigms (work domains, humans, technology). When applied to driver vigilance, an assessment would be considered *meaningless* (i.e., within a *meaning* processing account), without consideration of the concurrent contextual aspects surrounding the assessment of the driver. The recently introduced driver attention theory dubbed '*Minimum Required Attention*', proposes that '*a driver should only be considered inattentive when information sampling is not sufficient*' to the demands of the situation, '*regardless of whether the driver is executing an additional task or not*' (Kircher & Ahlstrom, 2017). In other words, observed behavior alone is not enough for assessments of distraction until placed in relation to situational/system demands. For example, typing a text message into a mobile phone while merging onto a busy commute highway connotes a different meaningful assessment of vigilance than the same actions while stopped at a red light in a rural town.

As opposed to presumed fixed-limits resource theories (cf. Wickens 1984, 1992), Malleable Attentional Resources Theory (Young & Stanton, 2002) has asserted that human attentional capacity naturally varies as a function of situational task demands (i.e., mental workload). In other words and observed by Hancock (2017):

'As the preeminent global adaptive species, humans readily learn and change their behavior in accordance with the constraints and opportunities of their ambient environment ... When we create boring, marginal, uninvolving interfaces to uninteresting tasks, we design boring, marginal, uninvolved and uninterested people. We cannot, in all good judgment, simply machine the mind to mind the machine.'

Thus, more consideration and measurement of the situations surrounding the driver are warranted to relate with those tools aimed with a human focus, as well as to design more meaningful interfaces to those relations.

Ecological approaches to driving safety can be traced to seminal work of Gibson and Crooks (1938). Their principles are re-advocated recently by Delucia and Jones (2017) such as: that organism-environment relations are the proper unit of analysis, that perception and action are continuous and cyclic, and that natural human perception is of relational affordances rather than object properties, etc. Situationally adaptive and appropriated understandings of driver distraction issues are no different. A technical task force of expertise from both European and US intelligent transportation systems researchers published a recent conceptual framework and taxonomy (Engstrom et al., 2013) that proposed a situated action-oriented view of attention and conceptualized driver inattention as '*mismatches between the driver's current resource allocation and that demanded by activities critical for safe driving, rather than in terms of attentional failures of the driver*'. Within Engstrom et al (2013), attentional allocations are viewed as adaptive functional processes regulating balances between benefits and costs where compensatory behavior emerges in regards to contexts (e.g., more attention in anticipation of demanding or uncertain situations such as complex intersections, focusing on detecting vehicles potentially appearing behind a blind corner, and/or uptake of non-driving activities when bored and/or sleepy).

4.2 DMS application fit

The present thesis provides results regarding relatable demands from specific driving scene features of road curvature (lateral course conflicts) and/or traffic volume (longitudinal collision conflicts) with specific eye measurements of movement (saccades, eccentricity, etc.). These results suggest a lower-level target for DMS applications to support the foundational core monitoring activity of driving: attentiveness in visuomotor control (whether of oneself under conventional driving circumstances or of another entity as with supervision over driving automation).

As discussed above, the driving task is clearly seen to be more than one thing, and monitoring activity (i.e., selective information input to action output mappings) is pervasive throughout. In formal research and engineering terms, hierarchical models are often employed to decompose/describe human driving performance along a framework of relational orderings. To clarify where the present thesis conclusions fit in and to what implications, it is helpful first to briefly overview a few of such models from Rasmussen (1983) and Parasuraman et al. (2000) for human cognitive information processing performance in general, and from Michon (1985), Merat et al. (2018), and Victor (2005) for driving problems specifically. The resulting overarching theme is one of consensus recognition and treatment of information that flows both fully and/or partially (semi-independently) through earlier/lower/faster and later/higher/slower mechanisms as typically mediated by experience/familiarity.

For all kinds of human operator performance and man-machine interface system designs, Rasmussen (1983) arranged what is known as the SRK framework with '*skill-based behavior*' (SBB) on the bottom, followed by '*rule-based behavior*' (RBB) next, and with '*knowledge-based behavior*' (KBB) on top. SBB involves a direct mapping between sensory input feature forms to automated sensori-motor pattern action outputs. At the RBB level, information proceeds through stages of recognition, association and rule retrieval as intercedents between sensory inputs and action outputs. At the KBB level, further intermediaries between sensory inputs and action outputs include identification, decisions, and planning.

Parasuraman et al (2000) adopted a simple four-stage view of human information processing proceeding in turn first from sensory processing, to perception/working memory, to decision making, and ultimately to response selection. In terms for modeling functions of automation they described these successive stages as information acquisition, information analysis, decision selection, and action implementation. Notably, they described such stages as capable of being considered as coordinated together in '*perception-action*' cycles (e.g., Gibson's (1979) affordance relations) rather than always in a strict serial sequence from stimulus to response. Similar accounts of information flow, with similar discussion of digressions off a singular path, are seminally represented by Endsley's (1995) three levels of situation awareness (perception, comprehension, and projection) as well as the how-what-why triads within abstraction hierarchies of cognitive work analysis from Vicente (1999) as explained within McIlroy & Stanton (2011) and in particular, the notional movement and shortcuts for information in control tasks across a '*decision ladder*' where:

'... although the diagram displays information processing in a linear fashion, different actors are likely to take different routes from the entry point to the end point. More specifically, novice workers are expected to follow the linear sequence while expert actors are often able to take shortcuts. For example, in certain situations the diagnosis of the system state may lead directly to the execution of a set procedure' (p. 363).

For driving specifically, Michon (1985) represents the problem solving tasks with three levels of skill and control that build cognitively upwards as a nested hierarchy with operational (control) on the bottom, tactical (maneuvering) in the middle, and lastly strategical (planning) on top. At the lowest level, environmental inputs are processed directly into automatic action pattern outputs in the timeframe of milliseconds, e.g., for threat coping aims to avoid acute, perceived danger consisting of the basic handling skills of steering and braking. In the middle, maneuvers produce controlled action patterns on the order of seconds, e.g., to negotiate merges, turns, and overtaking. At the top, strategies invoke general plans under a longer time constant, e.g., overall trip goals, route and modal choices.

Merat et al. (2018) adopted and extended the model of Michon (1985), merging it together with the levels of driving automation from SAE (2016), specifically for the conceptualization of '*out-of-the-loop*' in the automated driving problem domain. Therein a multi-level control of driving is depicted with continuous (ms – s), intermittent (s – min), and infrequent (min. – hrs.) monitoring activity inherent across driving control. For the innermost loop, the monitoring of lateral/longitudinal movements is tied in as '*basic vehicle motion control*' e.g., '*prediction of the movement of one's vehicle relative to other vehicles and within the lane ahead*'.

In his doctoral dissertation on roadway inattention, Victor (2005) emphasizes the criticality of the active vision approach that was argued to be '*relatively unknown to traffic researchers and human factors specialists developing in-vehicle information and communication systems and advanced driver assistance systems*' (p. 10). He introduced and explained a guiding principle of vision from Ungerleider and Mishkin (1982) consisting of two semi-independent cortical streams with foveal ocular time-sharing constraints: '*vision-for-action*' and '*vision-for-identification*' as in accordance with faster/basic ventral-stream processing compared to slower/abstracted dorsal-stream processing (see also '*System 1*' and '*System 2*' respectively in Kahneman, 2011). Such parallel division of labor of attentional processing allows for human drivers to move their eyes/attention both for fast and spatially accurate processing as visuomotor action control (e.g., immediate lateral and longitudinal protections) while at the same time for more conscious, representational, and goal-setting purposes (e.g., reading road signs and monitoring in-vehicle displays, etc). When coordination breaks down through competition for resources, the disruption of the lower attentional mechanism is explained via over-taxation from the higher attentional mechanism. In other words, thinking too heavily on a higher level (e.g., fixating too long on identifying/classifying an object or concept) detracts from vision-for-action loops of path- and headway-control.

The results and conclusions from the present thesis studies showed that not only are driving situations important for DMS assessments, but also suggest the level at which DMS could be promisingly targeted. The results evidenced that eye movement measurements (i.e., saccade amplitude, eccentricity, off-road glances) can be (beneficially) related to specific visual demands (i.e., amount of road curvature, amount of traffic). These measurements, of both eyes and scenes, reflect aspects of lateral and longitudinal spatial motion management — described by Michon (1985) and Merat et al. (2018) as operational functions/control and explained by Victor (2005) in terms of vision-for-action. **Thus, on the whole, the present thesis studies suggest means for DMS to be targeted to protect and maintain the lower foundational level or inner-most loop of driving attention (rather than interactive implicit layers and representational experiences that can be added on top).** The situated DMS affords an ability for the intelligent vehicle to be more judicious in its assessments and to conservatively refrain from alerting/reacting to simply whenever any

'secondary' higher cognitive tasking is presently detected (e.g., eye movement consequences of text reading or phone conversations, etc.). Instead, more (situationally) meaningful behavioral-based causes for caution/correction comes from situationally restricting DMS involvement to whenever the (measurable) eyes cannot (measurably) keep up with the (measurable) present visual demands for the most basic level of driving: safe/critical lateral and longitudinal vehicular control.

The power of the situated DMS comes from its protective involvement when a driver's thoughts have been decoupled from actions to such an extent that the (vision-for-action) eyes consequently suffer to keep up their basic lower level visuomotor control tasks of moving enough to match the visual demands of the present driving scene. In particular, because supervision of driving automation artificially splits the naturally adaptive perception-action cycle (i.e., Neisser, 1976) by asking for driving control perceptions from the human without his/her control actions, needs for DMS support are expected to be greater at such a level. As a composite homeostatic biological system, the vision-for-identification neural streams in the brain might be expected to overly dominate foveal occupations with diminished rehearsal requests from the vision-for-action neural streams which themselves then should reasonably carry metabolic and temporal '*start-up*' costs upon recall from periods of inactivity (like putting force on a muscle that has '*fallen asleep*' after sustained disuse). Thus, the earlier adaptive visual attention activity that is pre-cognitive in the sense that it sits before/below comprehension/awareness is expected to be where the results of the presently devised situated DMS might best fit in.

5. Future Research and Recommendations

The present thesis studies also provided new avenues in terms of automotive research methods.

Specifically, further descriptions and URLs for specific tools that were developed and felt potentially useful to future researchers (but were not otherwise available from the publications themselves) have been included directly as appendices to chapters, 3.1, 3.2, and 4.2. and are presently discussed in terms of extensibility.

Because future researchers and designers will ultimately be afforded and/or limited by their own available resources, a range of ways to know driving eyes and to know driving scenes are provided as contributions from the work of this thesis. For example, Chap 3.2 provides not only a theoretical corrective feedback loop '*big picture*' but also implementable regression equations that establish quantifiable relations between how much workload and attention different high/low driving scenes might be expected to require. For extensibility purposes, the situated DMS integration with driving automation of Chap 4.2 was designed with standard UDP communication protocols that separated customizable DMS classification states from consequential driving automation control actions that also included an abstraction layer for definition of course and/or collision conflict (e.g., all of which might differ between various automotive suppliers or research projects). Thus, it can be concluded from this thesis taken as a whole, that to develop DMS of driving vigilance, not only are eye measurements (esp. of movement distances) and scene contents (esp. road curvatures and collision hazards) important factors but they are obtainable in practical ways for future research and development applications.

Recommendations for future research fall under two general categories: (1) greater fidelity/complexity in driving simulations (e.g., more traffic, intersections, and real-life secondary tasks should provide greater generalizability of naturalistic driver adaption to driving scene demands) and (2) greater instrumentation technology in on-road vehicles (e.g., better knowledge of the driving scene contents and eye movement behaviors with improved measurement capabilities). Specifically out on the road, the presently available forward facing driving scene cameras had relatively low resolution and moved whenever the head of the participant moved thus obscuring/degrading recorded visual inputs for automatic computerized content segmentation let alone accurate pixel area coverage calculations. In its present form, the computational resources available to our driving simulator visualizations struggled with the additional scenery, sporadic oncoming traffic, and more than a single lead vehicle programmed to follow a specific trajectory within even the short 2-3 minute duration scenario which included only continuous traffic flow characteristics: no stop signs, intersections, merges, or turns.

More generally, initiative for broader areas of innovative driving research resources used in this thesis are summarized below, including the use of dashcam driving videos, crowdsourcing, and parallel eye-tracking.

5.1 Dash cam driving video recordings

In the conduct of research in the driving domain, it is easy to take for granted just how diverse driving can be. People commonly relate concepts to their own experiences and so it is a natural fallacy to disproportionately represent roads and driving conditions that are most familiar and available from one's own driving history and personal environment. Additionally, driving video recordings are growing research resources that offer a hybrid of enhanced stimulus/behavioral fidelity towards on-road applications that also allow for laboratory levels of repeatability and control.

Thus, in attempting an ecological approach to assessing driver eyes in context, a foundational interest of this thesis involved probing readily available sources of what driving really looks like. Casual observations across the last few years show a fairly steady increase in the number of dash cam driving video recordings publically posted and shared on YouTube, now currently totaling in excess of 5 million (Figure 5.1). Notably, by appending search terms such as 'extreme' or 'fail', many unusual, and often times dangerous, recordings of driving situations can be exploited in controlled and repeatable ways for various research purposes (e.g., which visual scene precursors to a hazardous driving event would occupants of an automated/autonomous car notice both with and without various kinds of infotainment and/or control interfaces).

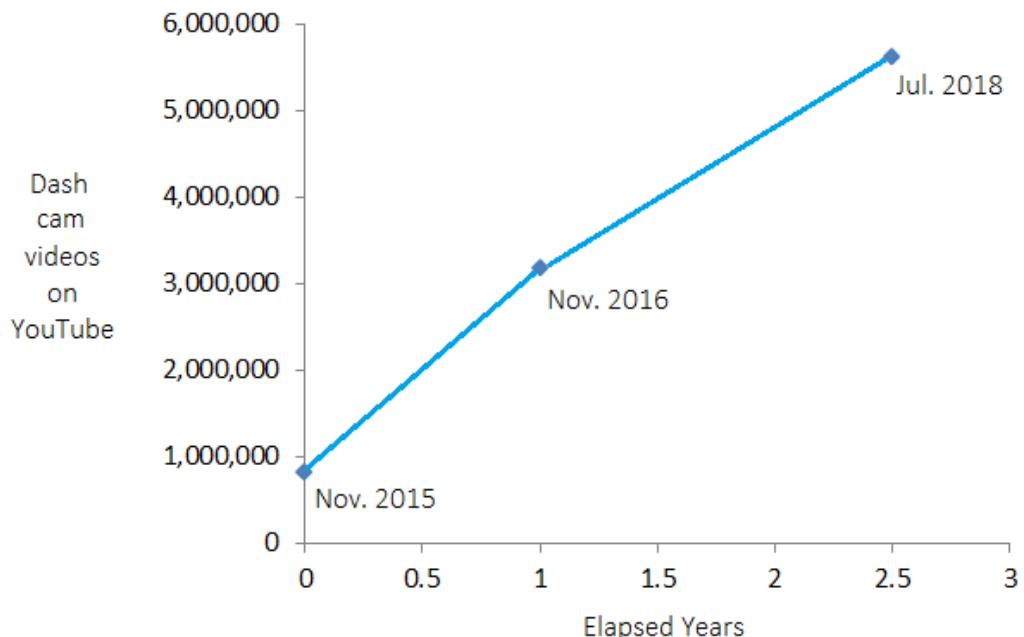


Figure 5.1. Recent growth in the number of publically available dash cam driving videos posted to YouTube

Furthermore, the automotive artificial intelligence (AI) community has long been actively contributing open source image data sets to advance the training and application of their machine learning computer vision models. Recently, dash cam driving video data sets have also been added to the community research pool. In June 2018, UC Berkeley teamed up with the AI dash cam company Nexar to release the BDD100K dataset containing 100,000 videos that include telemetry information like GPS locations, IMU data, and timestamps, as well as annotations such as object bounding boxes, lane marking identification, and indications of drivable areas.

5.2 Crowdsourcing for driving

The conventional driving transportation system has been built up over the last century by and for humans. From traffic engineers to city regulatory officials and drivers, humans design and consume the materials of the automotive system, so much in fact that driving skills (or knowledge of driving domain aspects) has become perhaps nearly on par with walking and talking. Meanwhile, across the world there is a growing community of online micro-task workers that through the Internet complete services in parallel with significant reductions in time and cost. Such crowd work continues to be a popular way of collecting subjective online survey data (esp. for aspects of innovative products not yet widely released such with various levels of driving automation). In the present chapter 3.1, objective work (i.e., naturalistic driving scene interpretation and content labeling) was newly explored and validated specifically in the traffic safety research domain. In a future of ever greater computer and car connectivity, it is conceivable that such a resource pool of in-common skills may further become useful to advance driving research and transportation service applications (e.g., tele-operated remote driving).

5.3 Parallel passenger eye-tracking comparisons

Much of the history of eye-tracking has had a lens focus towards measurements of a single individual, and across a pool of participants in separate sessions and typically in an enclosed laboratory environment. However, with massive recent reductions in cost, weight and size of camera and computing technology, eye-tracking equipment has evolved towards field studies and applications that involve the real world. Benefiting from the same goals and advances regarding reduced size and cost, the eye-tracking devices are increasingly prepared to undertake simultaneous measurements from multiple individuals. In the shared data collection methods of the present Chapter 4.1 and 4.2, comparisons of eye measures across a driving responsibility role of being in/out-of-the-loop were enabled in a real-world driving environment, with its many potential situational confounds of time, place, traffic conditions, weather, etc., held constant between the two roles of investigatory interest. With increasing driving automation, end-user research will correspondingly progress with topics involving non-driving vehicle participants (e.g., as everyone becomes a passenger interacting with an automated/autonomous driving agent). Simultaneous eye-tracking of multiple in-vehicle occupants allows for internal manipulations (e.g., different human machine interface designs) while controlling external conditions (all in the same vehicle at the same time and place) yet while retaining ecological real-world exposures of relevant driving scene situations.

6. Conclusion

Primary contributions of the present thesis regarding human factors of monitoring driving automation via eyes and scenes include: the critical **importance** of driving scene/situations (part 1); practically associated **measurement** constructs (part 2); and DMS-driving automation **integration** designs (part 3). Overall, it can thus be concluded that driver eyes adaptively move in relation to driving scene/situations and that details of both are measurable, such that situated DMS can be built and deployed with promising potential. Specifically, improved DMS are expected to improve human-automation interaction in terms of calibrated trust, enhanced acceptance, and more frequent and appropriate adherence. Consequently, road safety should reasonably be expected to increase and alleviate damaging societal costs.

References

Bainbridge, L. (1983). Ironies of automation. *Automatica*, vol. 19(6), pp. 775 - 779.

Banks, V.A., Eriksson, A., O'Donoghue, J., & Stanton, N.A. (2018). Is partially automated driving a bad idea? Observations from an on-road study. *Applied Ergonomics*, vol. 68, pp. 138 - 145.

Cichowski, J. (2017). Road warrior: Would you disconnect a safety feature? News, northjersey.com, USA Today Network. Available online (last accessed July 4, 2018) at: <https://www.northjersey.com/story/news/columnists/john-cichowski/2017/02/16/road-warrior-would-you-disconnect-safety-feature/97970160/>

Delucia, P.R. & Jones, K.S. (2017). Gibson and Crooks (1938): Vision and Validation. *The American Journal of Psychology*, vol. 130 (4), pp. 413 - 429.

Dolder, Falco, and Reese (2018). Autopilot Buddy ® patent pending Tesla Autopilot Nag Reduction Device. Available online (last accessed July 4, 2018) at: <https://www.autopilotbuddy.com/>

Endsley, M.A. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), pp. 32-54.

Endsley, M.R. (2017). Autonomous driving systems: A preliminary naturalistic study of the Tesla model S. *Journal of Cognitive Engineering and Decision Making*, vol. 11(3), pp. 225 - 238.

Engström, J., Monk, C.A., Hanowski, R.J., Horrey, W.J., Lee, J.D., McGehee, D.V., Regan, M., Stevens, A., Traube, E., Tuukkanen, M., Victor, T., & Yang, D. (2013). A conceptual framework and taxonomy for understanding and categorizing driver inattention. Brussels, Belgium: European Commission.

Gibson, J.J., (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton-Mifflin.

Gibson, J.J. & Crooks, L.E. (1938). A theoretical field-analysis of automobile-driving. *The American Journal of Psychology*, vol. 51, pp. 453 - 471.

Greenlee, E.T., DeLucia, P.R., & Newton, D.C. (2018). Driver vigilance in automated vehicles: Hazard detection failures are a matter of time. *Human Factors*, vol. 60(4), pp. 465 - 476.

Hancock, P.A. (2013). In search of vigilance: The problem of iatrogenically created psychological phenomena. *American Psychologist*, vol. 68(2), pp. 97 - 109.

Hancock, P.A. (2017). On the nature of vigilance. *Human Factors*, vol. 59(1), pp. 35 - 43.

Kahneman, D. (2011). *Thinking Fast and Slow*. New York, NY: Farrar, Straus, and Giroux.

Kircher, K. & Ahlstrom, C. (2017). Minimum Required Attention: A human-centered approach to driver inattention. *Human Factors*, vol. 59(3), pp. 471 - 484.

McIlroy, R.C., & Stanton, N.A. (2011). Getting past first base: Going all the way with Cognitive Work Analysis. *Applied Ergonomics*, 42, pp. 358-370.

Merat, N., Seppelt, B., Louw, T., Engstrom, J., Lee, J.D., Johansson, E., Green, C.A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., & Keinath, A. (2018). The 'out-of-the-loop' concept in automated driving: proposed definition, measures and implications. *Cognition, Technology & Work*. <https://doi.org/10.1007/s10111-018-0525-8>

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*. New York: Plenum Press, pp. 485 - 520.

Neisser, U. (1976). *Principles and Implications of Cognitive Psychology*. San Francisco, CA: W.H. Freeman

NSC (2017). Driver safety public opinion poll, February, 2017. National Safety Council. Available online (last accessed July 4, 2018) at: <http://d31hzlhk6di2h5.cloudfront.net/20170223/b7/89/d3/23/6d6608f44e99d10e54ec58e0/Driver-Safety-Poll.pdf>

NHTSA (2017). ODI Resume. Available online (last accessed July 2, 2018) at: <https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.PDF>

NTSB (2017). Collision between a car operating with automated vehicle control systems and a

tractor-semitrailer truck. National Transportation Safety Board, public meeting, NTSB/HAE 17-XX. Available online (last accessed July 4, 2018) at:
<https://www.ntsb.gov/news/events/Documents/2017-HWY16FH018-BMG-abstract.pdf>

Parasuraman, R., Sheridan, T.B., Wickens, C.D. (2000). Model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 30(3), pp. 286-297

Plungis, J. & Barry, K. (2018). Backup driver for Uber test car streamed Hulu at time of fatal crash, police say. Consumer Reports. June 22. Available online (last accessed July 4, 2018) at:
<https://www.consumerreports.org/autonomous-driving/backup-driver-for-uber-test-car-streamed-hulu-at-time-of-fatal-crash/>

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(3), pp. 257-266.

SAE (2016). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. In: SAE standard J3016 201609. SAE International, Warrendale

Stapel, J., Mullakkal-Babu, F.A., & Happee, R. (2019). Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving. *Transportation Research part F*, 60, pp. 590–605

Strauch, B. (2018). Ironies of Automation: Still unresolved after all these years. *IEEE Transactions on Human-Machine Systems*, 48(5), pp. 419-433

Stumpf, R. (2018). People keep coming up with ways to fool Tesla's Autopilot. The Drive. February 1, 2018. . Available online (last accessed July 4, 2018) at:
<http://www.thedrive.com/sheetmetal/18168/people-keep-coming-up-with-ways-to-fool-teslas-autopilot>

Underleider, L.G., & Mishkin, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale, & R.J.W. Mansfield (Eds.), *Analysis of Visual Behavior*, Cambridge, MA: MIT Press. pp. 549-586

Vicente, K.J. (1999). *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*. Mahwah, NJ, USA: Lawrence Erlbaum Associates

Victor, T. (2005). Keeping eye and mind on the road (Doctoral dissertation, Uppsala University, Uppsala, Sweden). Available from: <http://www.diva-portal.org/smash/record.jsf?pid=diva2:167500>

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention*. New York: Academic. pp. 63 - 101.

Wickens, C.D. (1992). Engineering psychology and human performance (2nd ed.). New York: Longman.

Young, M.S. & Stanton, N.A. (2002). Malleable Attentional Resources Theory: A new explanation for the effects of mental underload on performance. *Human Factors*, vol. 44(3), pp. 365 - 375.

Acknowledgments

A doctoral dissertation is an arduous journey with many people owed gratitude along the way.

First, I would like to thank my supervisory team. To Joost de Winter and Riender Happee, thank you for welcoming me into the auspicious and exciting Human Factors of Automated Driving consortium that you put together - it is a powerful European academic and professional traffic safety research network that I will always be proud with which to be associated. Your multiple leadership roles and committed work to the group as a whole and locally to our home contingency at Delft were very appreciated and inspiring. In regards to my personal development, I am forever grateful for your directed guidance and support throughout my many different literature reviews and experimental efforts. Joost, you have a truly unique gift for tracking an unparalleled volume of relevant literature references. Riender, thank you for helping me distill my work through productive prioritization. As my originating formal promoter, Frans van der Helm, thank you for allowing me to undertake these studies under your ultimate responsibility. You have startled me before, when with closed eyes you look as if to be dozing during our discussion sessions, but always in turn bring the conversation to an important point that otherwise would be missed. In the cold Dutch winters and amidst unfamiliar customs, you are a wise and warm Sinterklaas figure especially appreciated by this international alien.

Although, not directly under their tutelage, I also wish to express thanks to other leaders and instructors both locally within in the departments I inhabited as well as those I frequently visited. First mention is deserved to Dimitra Dodou both for my multiple opportunities to lead bachelor student thesis projects (along topics of my own choosing) as well as her tireless and extensive '*outside*' readings and critical feedback used to strengthen so much of my written work. Next, I am highly grateful for the positivity and good-natured energy of the likes of Dariu Gavrilla and David Abbink. With a rare combination of inter-personal character and command of exciting research areas, I was highly motivated to check for email announcements from them for the many interesting guest lectures and events that only they could thus draw in to the benefit of everyone nearby them. To those on the other side of campus, I am very grateful to Max Mulder, Renee Van Paassen, and Clark Borst who may not have always known I was eagerly auditing various of their aerospace lectures for their engaging perspectives on ecological and cognitive systems engineering approaches towards human-machine interaction.

To my cohort of PhD candidates, misery loves company and your company was well appreciated throughout both the ups and downs of being a PhD student.

Across the HFAuto consortium, thank you to all my fellow HFAuto early stage researchers (ESRs) and especially to our collective big brother experienced researcher (ER), Miltos Kyriakidis for his unique guidance and perspective. Thank you to Daniel Heikoop for your whimsy (and esp. your work to organize a proliferate of cognitive constructs). Thank you to Zhenji Lu for keeping words short, and to Matt Sassman for keeping words long. Thank you to Alexander Eriksson for your theoretical discussions and convictions. Thank you to Veronika Petrovych for actually bringing our drivenger study to life. Thank you to Bastiaan Petermeijer for introducing me to the glory of Hudito.

Thank you to Pavlo Bazilinsky for your enduring and endearing gumption. Thank you to Joel Goncalves for inspiring software architecture ambitions. Thank you to Alberto Morando for your extensive early devotion to road-safety annotation work, and to insights I am sure only you could be prepared to bear upon the domain. Thank you to Bo Zhang, the combined substance and style of your presentations were my personal favorite at our consortium meetings. Thank you to Silvia Varotto for your harsh critiques of my own style of presentation (and for your friendship, coziness and generous hospitality). Thank you to Ignacio Solis for your charm, as well as your academic rigor.

To those PhD comrades locally at TU Delft (both within and outside of HFAuto) thank you for our many lunch-table and hallway discussions. To Jork Stapel, thank you for never giving up on anything and for fielding countless questions from me about how to do something I want in Matlab and/or Simulink. To Felix Dreger, thank you for your closeness and expertise to cognitive psychology topics that are too easily forgotten or never directly encountered with such depth. Thank you to Natalia Kovacsova for the exciting prospect, even if never realized promise, to ride your motion-based motorcycle simulator and same to you Eline van der Kruk, regarding your instrumented ice-hockey skates. Thank you to Oliver Lee for being a piece of my hometown SF Bay Area that I could find in the room next door (and Tricia Gibo downstairs) rather than half the world away. Thank you to George Dialynas for always being able to put a smile on my face. Apologies are owed to Barys Shyrokau for temporarily misplacing the power adapter for the Logitech steering wheel you let me borrow, and a great deal of thanks for becoming a dear source of friendship and someone to look up to.

A great deal of gratitude is owed to all the Master and Bachelor students whom I either had the opportunity to directly supervise (and intrude in their project work) or to present lectures to and evaluate assignments in their course work. Thank you for putting up with my verbose thoughts, many questions/challenges and (re)directions, and for helping me to grow as an instructor/leader. Thank you especially to Pavan Besemer, Koen Jongbloed, Niek Rutten, Mitchel Knipscheer, Bas Lottman, Philip Oomkens, and Jasper Schol for letting me guide/shape your Bachelor thesis projects into preliminary explorations in human remotely driven automobiles. Thank you to Gerhard Marquart, Nico Janssen, Sonja Schneider, Yke Eisma, and Jonas Pijnenburg for allowing me to supervise and learn along with each of you through your individual Master's thesis experiments.

I am deeply thankful for the love and encouragement of all my friends and family, whether near or remote. Thanks especially to Peter Zaal and Thomas Lombaerts for introducing TU Delft to me and supporting my international acclimation each step of the way from initial house searches to final graduation customs. Thank you to Piotr Tomasik for lending me his garage in Las Vegas to build driving camera head-rest mounts in order to film both busy streets and empty desert roads. Thank you to both my little and big brother for just being the same, despite all the changes over the last few years. Thank you to my Mom for still taking care of me even as I continue through every stage of my adulthood. Thank you Mom, especially for diligently sending me every interesting piece of news you found regarding my timely topic of study; many of them were more relevant and useful to me than I might have let on at the time. I am deeply fortunate for the opportunity to thank Daphne Rein-Weston, my beloved Pilot, for guiding me and keeping me centered throughout (before and after) this endeavor. And to my dear dog Bentley, thank you for sharing your invaluable and formative puppy months with a PhD candidate who needs to get through just another few lines of thesis revisions before coming downstairs to play fetch with you in the backyard and to tell you that you are a good dog.

Lastly, thank you in advance to the future members of my thesis committee. Thank you for volunteering time out of your already busy professional/personal lives to read through my thesis work. I am greatly appreciative of what future discussions we will have and what challenges you will present to me to defend my readiness to obtain the esteemed rank you yourselves have previously acquired. In respects to the Prometheus mascot of the Delft University of Technology, I thank you and look forward for the opportunity to continue to carry the torch of knowledge forward.

List of Publications

Theses

Cabral, C.D.D. (2019).

Human Factors of Monitoring Driving Automation: Eyes and Scenes.

Doctoral dissertation, Delft University of Technology. Delft, South Holland, the Netherlands.

Cabral, C.D.D. (2010).

Aircraft deconfliction responsibility across en route sectors in NextGen separation assurance.

Master's thesis, San Jose State University. San Jose, California, USA.

Journal publications

Cabral, C.D.D., Stapel, J.C.J., Happee, R., & de Winter, J.C.F. (under review).

Redesigning today's driving automation towards adaptive backup control with situated and implicit interfaces.

Cabral, C.D.D., Petrovych, V., de Winter, J.C.F. (under review).

On-road driver vs. passenger eye eccentricity in a conventional car for in- vs. out-of-the-loop "drivenger" monitoring in automated vehicles.

Cabral, C.D.D., Happee, R., & de Winter, J.C.F. (under review).

Estimating driver readiness from situated eye movements: Prediction of workload and attention requirements from quantification of driving scene components.

Cabral, C.D.D., Eriksson, A., Dreger, F., Happee, R., & de Winter, J.C.F. (2019).

How to keep drivers engaged while supervising driving automation? A literature survey and categorization of six solution areas. *Driving Automation and Autonomy: Thematic issue of Theoretical Issues in Ergonomics (TIES)*, vol. 20(3), pp. 332-365
<https://doi.org/10.1080/1463922X.2018.1528484>

*Eisma, Y.B., *de Winter, J.C.F., Cabral, C.D.D., Hancock, P.A., & Stanton, N.A. (2018).

Situation awareness based on eye movements in relation to the task environment. *Cognition, Technology, & Work*, <https://doi.org/10.1007/s10111-018-0527-6> *Joint first authors

*Eisma, Y.B., *Cabral, C.D.D., & de Winter, J.C.F. (2018).

Visual sampling processes revisited: Replicating and extending Senders (1983) using modern eye-tracking equipment. *IEEE Transactions on Human-Machine Systems*, vol. 48(5), pp. 529-540, <https://doi.org/10.1109/THMS.2018.2806200> *Joint first authors

*Cabral, C.D.D., *Janssen, N.M., & de Winter J.C.F. (2018).

Adaptive automation: automatically (dis)engaging automation during visually distracted driving. *PeerJ Computer Science*, 4:e166, <https://doi.org/10.7717/peerj-cs.166> *Joint first authors

Kovacsova, N., Cabral, C.D.D., Antonisse, S.J., de Haan, T., van Namen, R., Nooren, J.L., Schreurs, R., Hagenzieker, M.P., & de Winter, J.C.F. (2018).

Cyclists' eye movements and crossing judgments at uncontrolled intersections: An eye-tracking study using animated video clips. *Accident Analysis & Prevention*, vol. 120, pp. 270-280, <https://doi.org/10.1016/j.aap.2018.08.024>

Cabral, C.D.D., Lu, Z., Kyriakidis, M., Manca, L., Dijksterhuis, C., Happee, R., & de Winter, J. (2018).

Validity and reliability of naturalistic driving scene categorization judgments from crowdsourcing. *Accident Analysis & Prevention*, vol. 114, pp. 25-33, <https://doi.org/10.1016/j.aap.2017.08.036>

Lu, Z., Happee, R., Cabral, C.D.D., Kyriakidis, M., & de Winter, J.C.F. (2016).

Human factors of transitions in automated driving: A general framework and literature survey. *Transportation Research Part F: Traffic Psychology and Behavior*, vol. 43, pp. 183-198, <https://doi.org/10.1016/j.trf.2016.10.007>

Cabral, C.D.D., Happee, R., & de Winter, J.C.F. (2016).

From Mackworth's clock to the open road: A literature review on driver vigilance task operationalization. *Transportation Research Part F: Traffic Psychology and Behavior*, vol. 40, pgs. 169-189, <https://doi.org/10.1016/j.trf.2016.04.001>

Lee, P., Smith, N., Brasil, C., Chevalley, E., Homola, J., Parke, B., Yoo, H., Bienert, N., Borade, A., Buckley, N., Cabral, C.D.D., Omar, F., & Gabriel, C. (2015).

Reducing departure delays at LaGuardia airport with Departure-Sensitive Arrival Spacing (DSAS). *Air Traffic Control Quarterly*, vol. 23(4), pp. 245-273. <https://doi.org/10.2514/atcq.23.4.245>

Prevot, T., Homola, J.R., Martin, L.H., Mercer, J., & Cabral, C.D.D. (2012)

Towards automated air traffic control: Investigating a fundamental paradigm shift in human/systems interaction. *Special Issue on NextGen: International Journal of Human-Computer Interaction*, vol. 28(2), pp. 77-98. <https://doi.org/10.1080/10447318.2012.634756>

Conference proceedings

Cabral, C.D.D., Stapel, J.C.J., Besemer, P., Jongbloed, K., Knipscheer, M., Lottman, B., Oomkens, P., Rutten, N., & Schol, J. (2019).

Plausibility of human remote driving: Human-centered experiments from the point of view of teledrivers and telepassengers. *63rd International Annual Meeting of the Human Factors and Ergonomics Society, Seattle, Washington, USA*

Cabral, C.D.D., Sheridan, T.B., Prevot, T., de Winter, J.C.F., & Happee, R. (2018).

The 4D LINT model of function allocation: Spatial-temporal arrangement and levels of automation. Proceedings of the 1st International Conference on Intelligent Human Systems Integration, 07-09 Jan, Dubai, UAE. In: Karwowski, W., & Ahram, T. (eds.) *Advances in Intelligent Systems and Computing*, vol. 722, pp. 29-34 https://doi.org/10.1007/978-3-319-73888-8_6

Cabral, C.D.D., Eriksson, A., Lu, Z., Petermeijer, S.M. (2018).

Current insights in human factors of automated driving and future outlook towards tele-operated remote driving services. Proceedings of the 1st International Conference on Intelligent Human Systems Integration, 07-09 Jan, Dubai, UAE. In: Karwowski, W., & Ahram,

T. (eds.) *Advances in Intelligent Systems and Computing*, vol. 722, pp. 10-12
https://doi.org/10.1007/978-3-319-73888-8_2

Cabral, C.D.D., Petrovych, V., & Happee, R. (2018).

Looking at drivers and passengers to inform automated driver state monitoring of in and out of the loop. *Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics (AHFE)*, 17-21 Jul, Los Angeles, CA, USA. In: Stanton, N. (ed.) *Advances in Human Aspects of Transportation, Advances in Intelligent Systems and Computing*, vol. 597, pp. 695-707 https://doi.org/10.1007/978-3-319-60441-1_67.

Cabral, C.D.D., Goncalves, J., Morando, A., Sassman, M., de Winter, J.C.F., & Happee, R. (2017).

Human Factors of Automated Driving (HFAuto) Deliverable D3.1 Driver-state monitor.
<https://doi.org/10.13140/RG.2.2.28169.65126/1>

Cabral, C.D.D., Happee, R., de Winter, J.C.F. (2016).

Similarity and differences between driving scenes: Objective and subjective descriptions in support of automated driving interactions. Poster presented at: *Annual Meeting of the Human Factors and Ergonomics Society Europe Chapter (HFES Europe)*, 26-28 Oct, Czech Republic, Prague.

<http://www.hfes-europe.org/wp-content/uploads/2016/10/Cabral2016poster.pdf>

Cabral, C.D.D., Janssen, N., Goncalves, J., Morando, A., Sassman, M., & de Winter, J.C.F. (2016).

Eye-based driver state monitor of distraction, drowsiness, and cognitive load for transitions of control in automated driving. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 09-12 Oct, Budapest, Hungary.
<https://doi.org/10.1109/SMC.2016.7844530>

Marquart, G., Cabral, C.D.D., & de Winter, J.C.F. (2015).

Review of eye-related measures of drivers' mental workload. *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics (AHFE)*, 26-30 Jul, Las Vegas, NV, USA. *Procedia Manufacturing*, vol. 3, pp. 2854-2861,
<https://doi.org/10.1016/j.promfg.2015.07.783>

Lee, P., Smith, N., Homola, J., Brasil, C., Buckley, N., Cabral, C.D.D., Chevalley, E., Parke, B., & Yoo, H. (2015).

Reducing departure delays at LaGuardia airport with Departure-Sensitive Arrival Spacing (DSAS). *Proceedings of the 11th USA/Europe Air Traffic Management Research and Development Seminar (ATM)*, paper 535, Eurocontrol/FAA. 23-26 Jun, Lisbon, Portugal.
http://www.atmseminar.org/seminarContent/seminar11/papers/535_Lee_0127150814-Final-Paper-5-4-15.pdf
http://www.atmseminar.org/seminarContent/seminar11/presentations/535-Lee_0127150814-PresentationPDF-6-29-15.pdf

Cabral, C.D.D., Gomez, A.N., Homola, J.R., Hunt, S.M., Martin, L.H., Mercer, J., & Prevot, T. (2014).

Transitioning Resolution Responsibility Between the Controller and Automation Team in Simulated NextGen Separation Assurance. *Proceedings of the 3rd Electronic Navigation Research Institute (ENRI) Workshop on Air Traffic Management(ATM)/Communication, Navigation and Surveillance (CNS) (EIWAC)*, 19-22 Feb 2013, Tokyo, Japan. In: ENRI (eds.) *Air Traffic Management and Systems, Lecture Notes in Electrical Engineering*, vol. 290, pp. 147-172, https://doi.org/10.1007/978-4-431-54475-3_9

Mercer, J., Homola, J.R., Cabrall, C.D.D., Martin, L.H., Morey, S.E., Gomez, A.N., & Prevot, T. (2014). **Human-automation cooperation for separation assurance in future NextGen environments.** *Proceedings of the 8th International Conference on Human-Computer Interaction in Aerospace (HCI-Aero)*, 30 Jul – 01 Aug, Santa Clara, CA, USA. <https://doi.org/10.1145/2669592.2669644>

Prevot, T., Callantine, T., Lee, P., Mercer, J., Homola, J., Martin, L., Brasil, C., & Cabrall, C.D.D. (2014). **An overview of current capabilities and research activities in the Airspace Operations Laboratory at NASA Ames Research Center.** *Proceedings of the 14th American Institute of Aeronautics and Astronautics (AIAA) Aviation Technology, Integration, and Operations Conference*, 16-20 Jun, Atlanta, GA, USA. <https://doi.org/10.2514/6.2014-2860>

Gomez, A.N., Martin, L.H., Homola, J.R., Morey, S.E., Cabrall, C.D.D., Mercer, J., & Prevot, T. (2013). **Investigating the complexity of transitioning separation assurance tools into NextGen air traffic control.** *Proceedings of the 32nd Digital Avionics Systems Conference (DASC)*, 6-10 Oct, Syracuse, NY, USA. <https://doi.org/10.1109/DASC.2013.6712529>

Homola, J.R., Morey, S.E., Cabrall, C.D.D., Martin, L.H., Mercer, J., & Prevot, T. (2013). **Analysis of interactive conflict resolution tool usage in a mixed equipage environment.** *Proceedings of the 29th American Institute of Aeronautics and Astronautics (AIAA) Guidance, Navigation and Control (GNC) Conference*, 19-22 Aug, Boston, MA, USA. <https://doi.org/10.2514/6.2013-4700>

Kupfer, M., Cabrall, C.D.D., Callantine, T.J., Homola, J.R., & Mercer, J. (2013). **Designing scenarios for controller-in-the-loop air traffic simulations.** *Proceedings of the American Institute of Aeronautics and Astronautics (AIAA) Modeling and Simulation Technologies (MST) Conference*, 19-22 Aug, Boston, MA, USA. <https://doi.org/10.2514/6.2013-5156>

Morey, S.E., Prevot, T., Mercer, J., Martin, L., Bienert, N., Cabrall, C.D.D., Hunt, S., & Kraut, J. (2013). **Controller strategies for automation tool use under varying levels of trajectory prediction uncertainty.** *Proceedings of the 13th American Institute of Aeronautics and Astronautics (AIAA) Aviation Technology, Integration, and Operations (ATIO) Conference*, 12-14 Aug, Los Angeles, CA, USA. <https://doi.org/10.2514/6.2013-4262>

Smith, N.M., Parke, B.K., Lee, P.U., Homola, J.R., Brasil, C.L., Buckley, N.J., Cabrall, C.D.D., Chevalley, E., Lin, C., Morey, S.E., Omar, F.G., Rein-Weston, D.F., & Yoo, H. (2013). **NextGen operations in a simulated NY area airspace.** *Proceedings of the 32nd Digital Avionics Systems Conference (DASC)*, 6-10 Oct, Syracuse, NY, USA. <https://doi.org/10.1109/DASC.2013.6712607>

Wing, D., Prevot, T., Lewis, T., Martin, L.H., Johnson, S., Cabrall, C.D.D., Commo, S., Homola, J.R., Sheth-Chandra, M., Mercer, J., & Morey, S.E. (2013). **Pilot and controller evaluations of separation function allocation in air traffic management.** *Proceedings of the 10th USA/Europe Air Traffic Management Research and Development Seminar (ATM)*, paper 299, Eurocontrol/FAA. 10-13 Jun, Chicago, IL, USA. http://www.atmseminar.org/seminarContent/seminar10/papers/299-Wing_0127130452-Final-Paper-4-12-13.pdf http://www.atmseminar.org/seminarContent/seminar10/presentations/299-Wing_0127130452-PresentationPDF-5-24-13.pdf

Cabral, C.D.D., Callantine, T.J., Kupfer, M., Martin, L.H., & Mercer, J. (2012). **Controller-managed spacing within mixed-equipage arrival operations involving Flight-Deck interval Management.** In Salvendy, G., & Karwowski, W. (eds.) *Proceedings of the 4th International Conference on Applied Human Factors and Ergonomics (AHFE)*, July 21-25, San Francisco, CA, USA. ISBN 9781466552623

Callantine, T.J., Cabral, C.D.D., Kupfer, M., Martin, L.H., Mercer, J. & Prevot, T. (2012). **Initial investigations of controller tools and procedures for schedule-based arrival operations with mixed Flight-deck Interval Management equipage.** *Proceedings of the 12th American Institute of Aeronautics and Astronautics (AIAA) Aviation Technology, Integration, and Operations (ATIO) Conference*, 17-19 Sep, Indianapolis, IN, USA. <https://doi.org/10.2514/6.2012-5673>

Homola, J.R., Lee, P.U., Lee, H., Brasil, C., Gregg, S., Mainini, M., Martin, L.H., Cabral, C.D.D., & Mercer, J. (2012). **Design and evaluation of a corridors-in-the-sky concept: The benefits and feasibility of adding highly structured routes to a mixed equipage environment.** *Proceedings of the 28th American Institute of Aeronautics and Astronautics (AIAA) Guidance, Navigation and Control (GNC) Conference*, 13-16 Aug, Minneapolis, MN, USA. <https://doi.org/10.2514/6.2012-4525>

Martin, L.H., Mercer, J., Callantine, T.J., Kupfer, M., & Cabral, C.D.D. (2012). **Air traffic controllers' strategies in the terminal area under off-nominal conditions.** In Salvendy, G., & Karwowski, W. (eds.) *Proceedings of the 4th International Conference on Applied Human Factors and Ergonomics*, July 21-25, San Francisco, CA, USA. ISBN 9781466552623

Gregg, S., Martin, L., Homola, J.R., Lee, P.U., Mercer, J., Brasil, C.L., Cabral, C.D.D., & Lee, H. (2012) **Shifts in air traffic controllers' situation awareness during high altitude mixed equipage operations.** *Proceedings of the 56th Human Factors and Ergonomics Society Annual Meeting*, 22-26 Oct, Boston, MA, USA. vol. 56(1), pp. 95-99 <https://doi.org/10.1177/1071181312561040>

Callantine, T.J., Cabral, C.D.D., Kupfer, M., Martin, L.H., & Mercer, J. (2011). **Investigating the impact of off-nominal events on high-density 'green' arrivals.** *Proceedings of the 30th Digital Avionics Systems Conference (DASC)*, 16-20 Oct, Seattle, WA, USA. <https://doi.org/10.1109/DASC.2011.6096001>

Homola, J.R., Martin, L.H., Mercer, J., Cabral, C.D.D., & Prevot, T. (2011). **Allocation of functions in a far-term air traffic control environment.** *Proceedings of the 30th Digital Avionics Systems Conference (DASC)*, 16-20 Oct, Seattle, WA, USA. <https://doi.org/10.1109/DASC.2011.6095964>

Prevot, T., Homola, J.R., Martin, L.H., Mercer, J., & Cabral, C.D.D. (2011). **Automated air traffic control operations with weather and time-constraints.** *Proceedings of the 9th Air Traffic Management Research and Development Seminar (ATM)*, paper 137, Eurocontrol/FAA. 14-17 Jun, Berlin, Germany. <http://www.atmseminar.org/seminarContent/seminar9/papers/137-Prevot-Final-Paper-4-15-11.pdf> <http://www.atmseminar.org/seminarContent/seminar9/presentations/137-Prevot-PresentationPDF-6-3-11.pdf>

Prevot, T., Mercer, J., Martin, L.H., Homola, J.R., Cabrall, C.D.D., & Brasil, C.L. (2011). **Evaluation of high density air traffic operations with automation for separation assurance, weather avoidance and schedule conformance.** *Proceedings of the 11th American Institute of Aeronautics and Astronautics (AIAA) Aviation Technology, Integration, and Operations (ATIO) Conference*, 20-22 Sep, Virginia Beach, VA, USA. <https://doi.org/10.2514/6.2011-6890>

Cabrall, C.D.D., Prevot, T., Homola, J.R., & Mercer, J. (2010). **Comparison of deconfliction responsibility procedures for adjacent en-route sectors in NextGen separation assurance.** *Proceedings of the 54th Human Factors and Ergonomics Society Annual Meeting*, 27 Sep – 01 Oct, San Francisco, CA, USA. vol. 54(1), pp. 16-20 <https://doi.org/10.1177/154193121005400105>

Homola, J.R., Prevot, T., Mercer, J., Brasil, C.L., Martin, L.H., & Cabrall, C.D.D. (2010). **A controller-in-the-loop simulation of ground-based separation assurance in a NextGen environment.** *Proceedings of the 2nd Electronic Navigation Research Institute (ENRI) Workshop on ATM/Communication, Navigation and Surveillance (CNS) (EiWAC)*, 11-12 Nov 2010, Tokyo, Japan. https://humansystems.arc.nasa.gov/publications/034_EiWAC2010_Jeffrey_Homola.pdf <https://www.enri.go.jp/eiwc/2010/pdf/prete/EN2010-034.pdf>

Lee, P.U., Sheridan, T.B., Poage, J., Martin, L.H., Cabrall, C.D.D., & Jobe, K. (2010). **Identification and characterization of key human performance issues and research in the Next Generation Air Transportation System (NextGen).** *NASA Contractor Report – 216390. Moffett Field, CA: NASA Ames Research Center.* <https://humansystems.arc.nasa.gov/publications/CR-2010-216390.pdf>

Martin, L.H., Cabrall, C.D.D., Lee, P.U., & Jobe, K.K. (2010). **Potential human factors issues in NextGen ATC concepts.** *Proceedings of the 3rd International Applied Human Factors and Ergonomics (AHFE) conference*, 17-20 Jul, Miami, FL, USA.

Prevot, T., Mercer, J., Martin, L.H., Homola, J.R., Cabrall, C.D.D., & Brasil, C.L. (2010). **Evaluation of NextGen air traffic control operations with automated separation assurance.** *Proceedings of the 6th International Conference on Human-Computer Interaction in Aerospace (HCI-Aero) Conference*, 03-05 Nov, Cape Canaveral, FL, USA. <https://humansystems.arc.nasa.gov/publications/HCI-Aero10-Prevot-34-2010-12-08.pdf>

Cabrall, C.D.D., Martin, L.H., Lee, P.U., & Jobe, K.K. (2009). **Exploration of human factors issues within a future separation assurance concept.** *Proceedings of the 28th Digital Avionics Systems Conference*, 23-29 Oct, Orlando, FL, USA. <https://doi.org/10.1109/DASC.2009.5347469>

Homola, J.R., Prevot, T., Mercer, J., Mainini, M., Cabrall, C.D.D. (2009). **Human/automaton response strategies in tactical conflict situations.** *Proceedings of the 28th Digital Avionics Systems Conference*, 23-29 Oct, Orlando, FL, USA. <https://doi.org/10.1109/DASC.2009.5347468>

Prevot, T., Homola, J.R., Mercer, J., Mainini, M., & Cabrall, C.D.D. (2009). **Initial evaluation of air/ground operations with ground-based automated separation assurance.** *Proceedings of the 8th USA/Europe Air Traffic Management Research and*

Development Seminar (ATM), paper 494, Eurocontrol/FAA. 29 Jun – 02 Jul, Napa, CA, USA.
http://www.atmseminarus.org/seminarContent/seminar8/papers/p_138_SP.pdf

Acknowledgement in Publications

Markos, S., & Pollard, J.K. (2015).

Passenger train emergency systems: Single-level commuter rail car egress experiments. U.S. Department of Transportation, Federal Railroad Administration. Final Report. DOT/FRA/ORD-15/04. DOT-VNTSC-FRA-12-05. <https://trid.trb.org/view/1350375>

Rottman, B.M., Ahn, W.K., Sanislow, C., & Kim, N.S. (2009).

Can clinicians recognize DSM-IV personality disorders from five factor model descriptions of patient cases? *The American Journal of Psychiatry*, vol. 166(4), pp. 427-433
<https://doi.org/10.1176/appi.ajp.2008.08070972>

Chase, S., & Donohoe, C. (2007).

Constructing a low-cost ground vehicle driving simulator at an airport. U.S. Department of Transportation, Volpe National Transportation Systems Center. Final Report. DOT/FAA/AR-07/59. DOT-VNTSC-FAA-07-10. Office of Runway Safety and Operational Services. CreateSpace Independent Publishing Platform, ISBN-10: 1494314401

Curriculum Vitae

Personal

1984 April 22

Born in Berkeley, California, USA

Christopher.Cabral.02@gmail.com

<https://www.linkedin.com/in/christophercabrall/>

Education

2002 - 2004 Double Major in Psychology and Literature – Creative Writing
Transferred before graduation

2004 - 2007 Bachelor of Arts: Psychology and Linguistics, minor: Computer Science, Northeastern University, Boston, MA, USA

2014 - 2019 PhD: Human Factors Engineering, Delft University of Technology, Delft, South Holland, the Netherlands

Professional Experience

2005 - 2007 **Engineering Psychologist**
Human Factors Division, Volpe National Transportation
Systems Center, Cambridge, MA, USA

2007 - 2008 **Research Assistant/Associate**, Thinking Lab of Dr. Woo-kyoung Ahn, Department of Psychology, Yale University, New Haven, CT, USA

2008 - 2014 **Senior Research Psychologist**, Airspace Operations Lab
Human Systems Integration, SJSURF, NASA Ames Research Center, Moffett Field, CA, USA

2014 - 2019 **Marie Curie Research Fellow**, Human Factors of Automated Driving, BioMechanical Engineering, Cognitive Robotics, Intelligent Vehicles, Delft University of Technology
Delft, South Holland, the Netherlands

May-Jul 2016 Swedish National Road and Transport Research Institute (VTI)
3 mo. internship Linkoping, Sweden

Propositions Belonging to the PhD Thesis

These propositions are regarded as lending themselves to opposition and as defendable, and have been approved as such by the promotors prof. dr. F.C.T. van der Helm, dr. ir. J.C.F. de Winter, and dr. ir. R. Happée.

- (1) Human remote driving is safer than automated driving, and easier to achieve than autonomous driving.
- (2) Tesla set an irresponsible and unethical precedent by using end consumers to beta test their 'Autopilot' advanced driving assistance system on public roads.
- (3) The full replacement of human driving control with autonomous processes is an inappropriate aim for driving problems caused by human attentional errors.
- (4) Vigilance decrements occur more often in automated driving than in conventional driving.

This proposition pertains to this dissertation (Chaps. 2.1)

- (5) Eye tracking measures now enable adaptive transitions of driving control where backing up the human with automated driving control is safer than forcing a return to manual control.

This proposition pertains to this dissertation (Chaps. 4.1, 4.2)

- (6) For driver monitor systems, measurements of how people look around are stronger determinants of being 'in-the-loop' than where/what drivers look at.

This proposition pertains to this dissertation (Chaps. 3.2, 3.3)

- (7) Cognitivism has done a great disservice to applied human factors. Behaviorism deserves and is already mounting a come-back.

- (8) More focus is warranted on the first rather than second word in the ubiquitously adopted Situation Awareness construct.

- (9) Non-interactive real-life driving videos are under-realized transportation safety research resources that provide more generalizability than driving simulators and more control than on-road studies.

This proposition pertains to this dissertation (Chaps. 3.1, 3.2)

- (10) Human error is not something in need of being resolved

