

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/120235>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Metadata of the chapter that will be visualized in SpringerLink

Book Title	Human Interaction and Emerging Technologies	
Series Title		
Chapter Title	Integrating Trust in Automation into Driver State Monitoring Systems	
Copyright Year	2020	
Copyright HolderName	Springer Nature Switzerland AG	
Corresponding Author	Family Name	Perello-March
	Particle	
	Given Name	Jaume
	Prefix	
	Suffix	
	Role	
	Division	Warwick Manufacturing Group
	Organization	The University of Warwick
	Address	Coventry, UK
	Email	Jaume.Perello-March@warwick.ac.uk
Author	Family Name	Burns
	Particle	
	Given Name	Christopher
	Prefix	
	Suffix	
	Role	
	Division	Warwick Manufacturing Group
	Organization	The University of Warwick
	Address	Coventry, UK
	Email	C.Burns.2@warwick.ac.uk
Author	Family Name	Elliott
	Particle	
	Given Name	Mark
	Prefix	
	Suffix	
	Role	
	Division	Warwick Manufacturing Group
	Organization	The University of Warwick
	Address	Coventry, UK
	Email	M.T.Elliott@warwick.ac.uk
Author	Family Name	Birrell
	Particle	
	Given Name	Stewart
	Prefix	
	Suffix	

Role	
Division	Warwick Manufacturing Group
Organization	The University of Warwick
Address	Coventry, UK
Email	S.Birrell@warwick.ac.uk

Abstract

Inappropriate trust in highly automated vehicles (HAVs) has been identified as one of the causes in several accidents [1–3]. These accidents have evidenced the need to include a Driver State Monitoring System (DSMS) [4] into those HAVs which may require occasional manual driving. DSMS make use of several psychophysiological sensors to monitor the drivers' state, and have already been included in current production vehicles to detect drowsiness, fatigue and distractions [5]. However, DSMS have never been used to monitor Trust in Automation (TiA) states within HAVs yet. Based on recent findings, this paper proposes a new methodology to integrate TiA state-classification into DSMSs for future vehicles.

Keywords
(separated by '-')

Trust in Automation - Driver State - Monitoring Systems - Highly automated vehicles



Integrating Trust in Automation into Driver State Monitoring Systems

Jaume Perello-March^(✉), Christopher Burns, Mark Elliott,
and Stewart Birrell

Warwick Manufacturing Group, The University of Warwick, Coventry, UK
{Jaume.Perello-March, C.Burns.2, M.T.Elliott,
S.Birrell}@warwick.ac.uk

Abstract. Inappropriate trust in highly automated vehicles (HAVs) has been identified as one of the causes in several accidents [1–3]. These accidents have evidenced the need to include a Driver State Monitoring System (DSMS) [4] into those HAVs which may require occasional manual driving. DSMS make use of several psychophysiological sensors to monitor the drivers' state, and have already been included in current production vehicles to detect drowsiness, fatigue and distractions [5]. However, DSMS have never been used to monitor Trust in Automation (TiA) states within HAVs yet. Based on recent findings, this paper proposes a new methodology to integrate TiA state-classification into DSMSs for future vehicles.

Keywords: Trust in Automation · Driver State · Monitoring Systems · Highly automated vehicles

1 Introduction

Highly Automated Driving will change the manner in which we drive, will place new demands of their users, and as a consequence, new human factors such as TiA will arise. TiA is an attitude that leads to a behaviour - reliance - [6] from which several behavioural outcomes can result including correct use, misuse, disuse and overuse of automation [7]. In addition, TiA is a multidimensional construct [8], implying that TiA may have many states. The taxonomy proposed here for DSMS would classify them as Appropriate TiA, Over-TiA and Distrust in automation.

The current standardised methodology to evaluate TiA uses self-reporting tools [9, 10]. These tools have proven to be valid and feasible but are limited to experimental scenarios and post-hoc data analysis, and do not allow active real-time measurement. Therefore, a different methodology in accordance with DSMSs capabilities needs to be developed. Recent findings in the TiA literature have suggested a promising alternative for this problem – using psychophysiology [11, 12]. Using the existing TiA scales [9, 10], this review will focus on identifying the aforementioned TiA states and their psychophysiological correlates, and investigate the use of DSMS data to train a TiA classifier using machine learning algorithms.

2 TiA Decision-Making Processes

It has been proposed that TiA is an attitude involving both affective and cognitive decision-making processes [6, 13–15]. Following the proposal by Lee and See [6], these can be classified as:

- Analogical: trust judgements based on rules and heuristics, making use of previously known solutions for similar current problems, other individuals' experiences, societal norms, etc. Analogical decisions are less cognitively demanding than analytic judgements, allowing faster decision-making.
- Analytical: a judgement based on active evaluations of risks and benefits of trusting an automated system. It relies on knowledge about the automated system's characteristics and performance. Analytical decision-making is more cognitively demanding and time-consuming.
- Affective: This mechanism bases trust judgements on impressions, feelings and emotions regarding the trustee. Often users may not trust automation only because they feel uneasy, lacking any kind of reasoning. Affective processes are used in high time-pressured or when cognitive resources are not available to make a judgement [6].

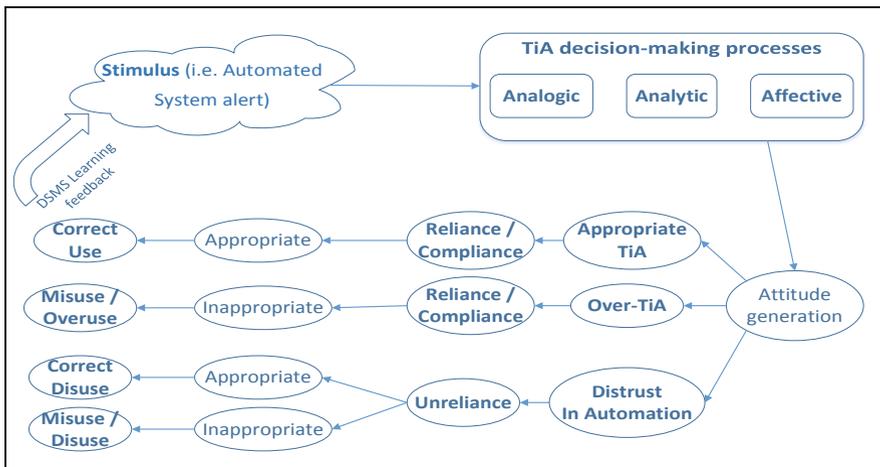


Fig. 1. Model of TiA process

The use of each mechanism depends on the cognitive resources available (knowledge about the system, similar past experiences, cognitive workload, etc.) and time pressures [6]. E.g. the user may rely on analytic processing when there is sufficient time and cognitive resources available but when cognitive resources are limited and time is constrained, the user may instead rely on faster, more subconscious analogical and affective processes [15]. Building upon the existing models of TiA, this paper proposes that attitudes generated may represent particular psychophysiological patterns, which importantly for DSMS could be evaluated in real-time. To enable future

research in this area this paper presents a diagrammatic representation of the initial decision-making, to final behavioural outcome, process (see Fig. 1). Importantly for our research this enables a controlled stimulus (or automated vehicle scenario) to be created, and physiological responses evaluated, leading to greater understanding of the TiA decision-making process which finalizes in an attitude being formed.

3 Modelling Trust in Automation

Previous authors have already proposed mathematical models to classify TiA [16, 17]. However these models are grounded on TiA frameworks [18, 19] which have been updated [6] and [15]. Recent developments in the Human-Robot interaction domain suggest that mathematically modelling TiA in HAVs based on mental workload and affective states is achievable [20, 21]. Therefore, it is worth considering updating and adapting these newer models to the present state-of-the-art for HAVs. The following section proposes a new methodology, building on Fig. 1, which could be used to classify TiA into three states using machine-learning algorithms:

1 - Appropriate Trust in Automation. An appropriate level of TiA means that the user's trust is calibrated accordingly with the automated system's reliability [22] – i.e. the user is aware of the system limitations and capabilities, and of the current traffic situation and uses the automated system appropriately. This is the desired state as it leads to a correct and safe use of the HAV. Appropriate TiA means that a user will be ready to take manual control when the systems requires by being aware that the system's capabilities are limited in certain situations [6]. Another example is that the user will rely upon and activate automated driving when suggested. Under such a scenario, it can be assumed that the user will be aware of the situation, engaged in the driving task and is confident about the automated system's capabilities. This raises the possibility of classifying appropriate TiA as an expected state of relaxation (i.e. positive valence and low arousal) with mental engagement. Previous studies have successfully classified relaxation states using Support Vector Machine (SVM), Regression Tree (RT), K-Nearest Neighbour (KNN) and Bayesian Network (BN) classifiers based on a Heart Rate Variability (HRV) decrease, an increase of Heart Rate (HR), Respiration Rate (RR) and Electro-Dermal Activity (EDA) [23, 24]. A similar method to that used in [24], could be also implemented for Electroencephalography (EEG), EDA and eye-tracking signals to classify emotional states. Mental engagement or increased mental workload has also been successfully classified using SVM [25, 26] and Artificial Neural Networks (ANN) [27–29]. These classifiers were based on an increase in pre-frontal activity via Event-Related Potentials from EEG, blood oxygenation increases from functional Near Infrared Spectroscopy (fNIRS), increased pupil size, fixations, saccades and reductions in blink-ratio. In addition, identification by Random Forest (RF) has been used for eye-tracking metrics [30].

2 - Over-Trust in Automation. This occurs when TiA is not properly calibrated due to e.g. higher expectations regarding the automated system's capabilities [6]. Over-trust is probably the most undesirable and hazardous state and some of the accidents

reported in the introduction are good examples [1, 2]. In this case, the user lacks knowledge regarding the actual system reliability and therefore tends to believe that the automated system will be able to appropriately contend with situations that it cannot. For example, users may not be expecting a request to take-back control and may miss or ignore this, as they are not aware of the current situation and the vehicle's performance. Therefore, over-TiA could be related to low arousal emotional states (from bored to sleepy) and a relative lack of cognitive engagement with the driving task which could be classified using SVM, RT, KNN and BN [23] based on reduced RR, EDA, HR and increased HRV. A lack of situational awareness, complacency, automation bias and even a state of drowsiness can be classified using Extreme Learning Machine (ELM) and SVM methods based on reduced pre-frontal activity via EEG or fNIRS. RF may also be used for eye-tracking parameters [30] such as increased blink ratios, reduced pupil size, reduced fixation ratio and saccadic rates [25].

3 - Distrust in Automation represents the state in which a user's self-confidence is higher than the expectations of system performance, and can be due to previous negative or inferior experiences with other, similar automated systems. Distrust may also be motivated by excessive system performance expectations that do not match the system's real capabilities, resulting in user dissatisfaction. Ultimately, the outcome of distrust usually leads to Automated System disuse [7]. Such as [3], where the HAV operator took manual control and turned right when approaching a leading vehicle, consequently colliding with an approaching motorbike on the right side. The HAV was about to reduce the speed when approaching the leading vehicle, and if the operator had not taken control, the crash would not have occurred. High-distrust situations could be associated with increased alertness or monitoring behaviours and can be classified using ELM [25], based on an increase on pre-frontal activity, pupil size, fixations, saccades and reduced blinking ratio, and RF for these visual parameters [30]. Increased mental workload can be classified using SVM [26] and ANN [27, 28], based on an increase on pre-frontal activity, increased HR, increase of tonic EDA responses, increased RR, increased pupil size, and reduced eye blinking ratio. Finally, affective states with high arousal and negative valence such as fear, distress or frustration can be expected and classified using SVM, RT, KNN and BN [23] based on increases in HR and RR, reduced HRV and increase in EDA as inputs.

4 Further Research and Conclusions

Logically, future research should focus on finding correlations between physiological data and TiA scales ([9, 10]) to identify basic TiA states (appropriate trust, distrust, and over-trust). The first study following up this paper will focus on test whether the expected psychophysiological tendencies suggested in Sect. 3 are confirmed. Automation reliability expectations will be generated on naïve HAVs users in order to build up different TiA among them. Using WMG's 3xD driving simulator, participants will be driven along scenarios of increasing complexity. We expect that their TiA will develop depending on the pre-administered system reliability expectations. We hypothesize that psychophysiological patterns associated with participants who trust

the HAV will be statistically different from those who distrust the system. These results may serve as a TiA baseline for further related studies and as a pool of data to feed the DSMS through machine learning techniques.

References

1. NTSB: Collision between a Car Operating with Automated Vehicle Control Systems and a Tractor-Semitrailer Truck. NTSB/HAR-17-XX (2016)
2. NTSB: Preliminary Report HWY18MH010, p. 4 (2018)
3. Waymo: The very human challenge of safe driving, 5 November 2018. <https://medium.com/waymo/the-very-human-challenge-of-safe-driving-58c4d2b4e8ee>. Accessed 10 Dec 2018
4. Melnicuk, V., Birrell, S., Crundall, E., Jennings, P.: Towards hybrid driver state monitoring: review, future perspectives and the role of consumer electronics. In: Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 1392–1397, August 2016
5. Ledezma-Zavala, E., Ramirez-Mendoza, R.A.: Towards a new framework for advanced driver assistance systems. *Int. J. Interact. Des. Manuf.* **12**(1), 215–223 (2018)
6. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum. Factors J. Hum. Factors Ergon. Soc.* **46**(1), 50–80 (2004)
7. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. *Hum. Factors* **39**(2), 230–253 (1997)
8. Spain, R.D., Bustamante, E.A., Bliss, J.P.: Towards an empirically developed scale for system trust: take two. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 52, no. 19, pp. 1335–1339 (2008)
9. Jian, J.-Y., Bisantz, A.M., Drury, C.: Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn. Ergon.* **4**(1), 53–71 (2000)
10. Körber, M.: Theoretical considerations and development of a questionnaire to measure trust in automation. In: 20th Triennial Congress of the IEA, pp. 1–20, March 2018
11. Morris, D.M., Erno, J.M., Pilcher, J.J.: Electrodermal response and automation trust during simulated self-driving car use. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 61, no. 1, pp. 1759–1762 (2017)
12. Waytz, A., Heafner, J., Epley, N.: The mind in the machine: anthropomorphism increases trust in an autonomous vehicle. *J. Exp. Soc. Psychol.* **52**, 113–117 (2014)
13. Jan De Visser, E.: *The World Is Not Enough: Trust in Cognitive Agents*. George Mason University, Fairfax, VA (2012)
14. Madsen, M., Gregor, S.: Measuring human-computer trust. In: Proceedings of the Eleventh Australasian Conference on Information Systems, pp. 6–8 (2000)
15. Hoff, K.A., Bashir, M.: Trust in automation: integrating empirical evidence on factors that influence trust. *Hum. Factors* **57**(3), 407–434 (2015)
16. Itoh, M., Tanaka, K.: Mathematical modeling of trust in automation: trust, distrust, and mistrust. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 1, no. 1994, p. 9 (2000)
17. Lee, J., Moray, N.: Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* **35**(10), 1243–1270 (1992)
18. Muir, B.M.: Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* **37**(11), 1905–1922 (1994)
19. Muir, B.M., Moray, N.: Trust in automation. Part ii. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* **39**(3), 429–460 (1996)

20. Wang, X., Shi, Z., Zhang, F., Wang, Y.: Mutual trust based scheduling for (semi) autonomous multi-agent systems. In: Proceedings of the American Control Conference, pp. 459–464, July 2015
21. Wang, Y., Shi, Z., Wang, C., Zhang, F.: Human-robot mutual trust in (semi)autonomous underwater robots. In: Koubaa, A., Khelil, A. (eds.) Cooperative Robots and Sensor Networks. Studies in Computational Intelligence, vol. 554, pp. 115–137. Springer, Heidelberg (2014)
22. Khastgir, S., Birrell, S., Dhadyalla, G., Jennings, P.: Calibrating trust through knowledge: introducing the concept of informed safety for automation in vehicles. *Transp. Res. Part C Emerg. Technol.* **96**(July), 290–303 (2018)
23. Rani, P., Liu, C., Sarkar, N.: An empirical study of machine learning techniques for affect recognition in human-robot interaction. In: 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, pp. 2451–2456 (2005)
24. López-Gil, J.-M., Virgili-Gomá, J., Gil, R., García, R.: Corrigendum: method for improving EEG based emotion recognition by combining it with synchronized biometric and eye tracking technologies in a non-invasive and low cost way. *Front. Comput. Neurosci.* **10** (November), 9–10 (2016)
25. Chen, L.L., Zhao, Y., Zhang, J., Zou, J.Z.: Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning. *Expert Syst. Appl.* **42**(21), 7344–7355 (2015)
26. Hogervorst, M.A., Brouwer, A., Van Erp, J.B.F.: Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. **8**, 1–14 (2014)
27. Wilson, G.F., Russell, C.A.: Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Hum. Factors J. Hum. Factors Ergon. Soc.* **45**(4), 635–644 (2003)
28. Wilson, G.F., Russell, C.A.: Performance enhancement in an uninhabited air vehicle task using psychophysiological determined adaptive aiding. *Hum. Factors* **49**(6), 1005–1018 (2007)
29. Wilson, G.F., Russell, C.A., Force, A., Base, A.F.: Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Hum. Factors* **45**(3), 381–389 (2003)
30. Zemblyns, R., Niehorster, D.C., Komogortsev, O., Holmqvist, K.: Using machine learning to detect events in eye-tracking data. *Behav. Res. Methods* **50**(1), 160–181 (2018)

Author Query Form

Book ID : **483905_1_En**

Chapter No : **53**

Please ensure you fill out your response to the queries raised below and return this form along with your corrections.

Dear Author,

During the process of typesetting your chapter, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the ‘Author’s response’ area provided below

Query Refs.	Details Required	Author’s Response
AQ1	Please provide full journal title for Ref. [26], if applicable.	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

<i>Instruction to printer</i>	<i>Textual mark</i>	<i>Marginal mark</i>
Leave unchanged	... under matter to remain	Ⓟ
Insert in text the matter indicated in the margin	⋈	New matter followed by ⋈ or ⋈ [Ⓢ]
Delete	/ through single character, rule or underline or ┌───┐ through all characters to be deleted	Ⓞ or Ⓞ [Ⓢ]
Substitute character or substitute part of one or more word(s)	/ through letter or ┌───┐ through characters	new character / or new characters /
Change to italics	— under matter to be changed	↙
Change to capitals	≡ under matter to be changed	≡
Change to small capitals	≡ under matter to be changed	≡
Change to bold type	~ under matter to be changed	~
Change to bold italic	≈ under matter to be changed	≈
Change to lower case	Encircle matter to be changed	≡
Change italic to upright type	(As above)	⊕
Change bold to non-bold type	(As above)	⊖
Insert 'superior' character	/ through character or ⋈ where required	Υ or Υ under character e.g. Υ or Υ
Insert 'inferior' character	(As above)	⋈ over character e.g. ⋈
Insert full stop	(As above)	⊙
Insert comma	(As above)	,
Insert single quotation marks	(As above)	ʹ or ʸ and/or ʹ or ʸ
Insert double quotation marks	(As above)	“ or ” and/or “ or ”
Insert hyphen	(As above)	⊞
Start new paragraph	┌	┌
No new paragraph	┐	┐
Transpose	└┐	└┐
Close up	linking ○ characters	⸸
Insert or substitute space between characters or words	/ through character or ⋈ where required	⸶
Reduce space between characters or words		⸵