Towards Acceptance of Automated Driving Systems

Samantha Jamson[1], Konstantinos Risvas[2], Roi Naveiro[3], David Ríos Insua[3],
Konstantinos Moustakas[2], Mikolaj Kruszewski[4], Aleksandra Rodak[4] and Alessandro Barisone[5]
[1]*Institute for Transport Studies, University of Leeds, U.K.*
[2]*Electrical and Computer Engineering Department, University of Patras, Greece*
[3]*Institute of Mathematical Sciences (ICMAT-CSIC), Madrid, Spain*
[4]*Motor Transport Institute, Warsaw, Poland*
[5]*algoWatt S.p.A, Italy*

Keywords: Automated Driving Systems, Trust and Acceptance, Request to Intervene, Decision Support, Human Machine Interfaces.

Abstract: The acceptance of Automated Driving Systems is of key importance since it will determine whether they will actually be used. This presentation describes contributions in this broad area from the perspective of the Trustonomy project with a focus on ethical decision support, human machine interfaces and trust assessment, aimed at enhancing the experience of drivers and passengers in such vehicles.

1 MOTIVATION

Automated driving systems (ADS) are poised to constitute a major technological innovation reshaping transportation as we know it. Recent breakthroughs in Artificial Intelligence (AI), coupled with advances in computational hardware, have had a revolutionary effect on ADS allowing cutting-edge control algorithms to be executed in real time. Despite these advances, it is generally recognised that ADS technology will not be widely deployed in the immediate future: its incorporation onto global roadways will be a gradual process (Mahmassani, 2016). Combined with the electrification of vehicles and a change in the concept of car ownership, ADS would conform a future in which we would expect fewer accidents, less pollution, less wasted travel time, and increased traveling possibilities for many collectives, including the elderly (Burns and Shulgan, 2019). This paper presents contributions from the Trustonomy project (`https://h2020-trustonomy.eu/`) in the areas of ethical decision support, human machine interfaces and human factors geared towards increasing trust and acceptance in ADS so as to accelerate their adoption.

2 CONTEXT

A gradual progression from manned vehicles (MV) to ADS is widely expected. Its stages are shown in the SAE six-level driving automation taxonomy, with level 0 describing vehicles with no automated capacity, and levels 1 through 5 representing vehicles with increasing automated features culminating in fully automated, level-5 ADS. Over the last decade, manufacturers have begun to produce vehicles of higher automation levels, in particular level-1 and -2. However, many crucial limitations related to ADS safety and operational robustness will likely restrict automated vehicles on roads to, at best, levels 3 and 4 over the next decade.

Figure 1 integrates relevant impacts of ADS through a diagram illustrating the interconnected nature of numerous factors related to their adoption. Dotted nodes refer to positive impacts; light grey nodes, to negative ones; white nodes, to other impacts, not necessarily positive or negative, that need to be addressed. Dark grey nodes refer to contextual factors (e.g., trust in ADS) that might have a major influence on the massive deployment of ADS.

Many factors in Figure 1 can be indeed readily assumed as positive. For example, provided that the driving error of human operators exceeds that

Jamson, S., Risvas, K., Naveiro, R., Insua, D., Moustakas, K., Kruszewski, M., Rodak, A. and Barisone, A.
Towards Acceptance of Automated Driving Systems.
DOI: 10.5220/0010721300003060
In *Proceedings of the 5th International Conference on Computer-Human Interaction Research and Applications (CHIRA 2021)*, pages 232-239
ISBN: 978-989-758-538-8; ISSN: 2184-3244

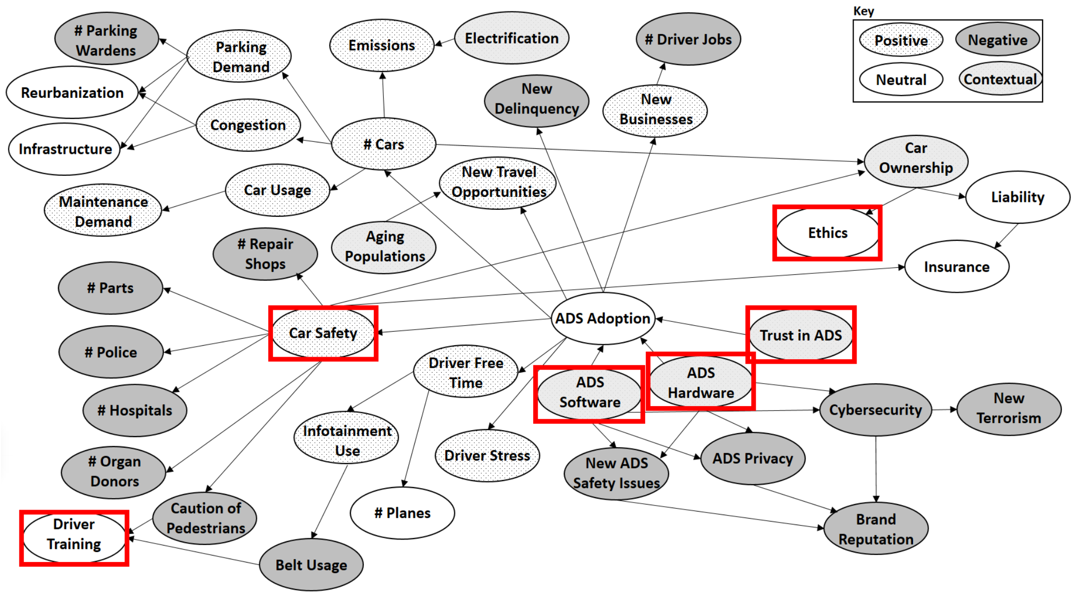


Figure 1: ADS Impacts on Society. Adapted from (Caballero et al., 2021b).

of ADS technology, roadway safety can be expected to increase. However, other impacts are perceived negatively: although some professions may be re-envisioned using ADS, others may be at risk, like e.g., taxi drivers, as a consequence of the availability of competitive automated taxis. Similarly, the elimination of human error does not imply the elimination of machine error. As a consequence, third parties manufacturing ADS safety systems will face greater vulnerability to liability lawsuits and reputation risk. Other impacts associated with the massive adoption of ADS will require a major re-definition of the current status quo, without necessarily having positive or negative connotations. For example, driver training will need to evolve.

Despite the major benefits that massive adoption of ADS will bring, a core issue with these new technologies is the need to build acceptance and trust in society to facilitate ADS adoption. This is the core objective of the Trustonomy project which drives our discussion here. Trustonomy involves 16 partners across Europe covering 4 pilots in Poland/Finland, UK, Italy and France. It is framed according to six pillars related with: (1) driver state monitoring (DSM) systems assessment; (2) curricula for driver training in ADS; (3) driver intervention performance assessment (DIPA); (4) ethical automated decision support framework, covering liability and risk assessment; (5) Human Machine Interfaces (HMI) assessment; and (6) measuring trust and acceptance. The related nodes in Figure 1 are marked with a red rectangle. For space reasons, we focus on the last three topics.

3 ETHICAL DECISION SUPPORT FOR ADS

For reasons outlined above, level -3 and -4 vehicles are the primary focus of Trustonomy. Such ADS require human intervention when operating outside of their specified operational design domain (ODD). That input is solicited through an HMI via a *request to intervene* (RtI) operation, which is core in the project. Until level-5 vehicles predominantly populate global roadways, RtI decisions and their management will remain a crucial, safety-related issue. Relatively few studies focus on the RtI management. This section sketches a solution.

Consider a level-3 or -4 ADS for which several driving modes are available (typically, *automated*, *manual* and *emergency*). A decision-analytic framework is utilized to manage the RtI operations as in Figure 2. The main workflow can be summarized as follows. At each time t, the environmental and DSM systems observe both the state of the environment and the driver. Based on past observations, a forecast of future states is produced and the planned trajectory is updated. If the forecasts indicate that it is very likely that the vehicle will be outside its ODD limits in the near future, an RtI should be executed. If the RtI is accepted, the intervention will be assessed via a DIPA. This approach emphasizes a *management by exception principle* (West and Harrison, 2006) wherein a group of models is used for inference, prediction and decision support under standard

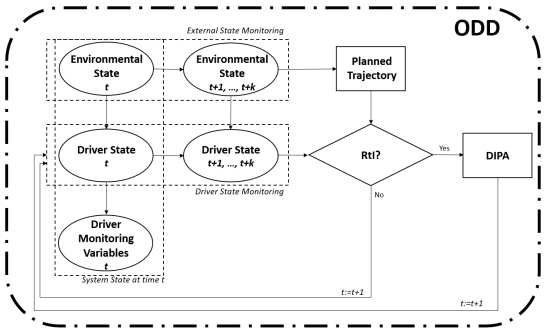


Figure 2: Decision support for ADS.

driving circumstances until an exception arises that triggers an RtI. The approach incorporates warnings to raise driver awareness which can be modulated in two directions: (1) several alert levels (e.g., warning and critical) can be introduced; (2) if an alert must be issued repeatedly, the HMI system can successively amplify the alert (e.g., increase volume of subsequent warnings).

To integrate and update information from key sources in a coherent manner, a Bayesian approach is utilized based upon observed system behavior. D_t will designate the data set available up to time t incorporated from the ADS sensors. Typically, most ADS maneuvers, tasks and local trajectory planning are scheduled a few steps ahead, e.g., $k = 10$ time intervals of 0.5 seconds, depending on driving conditions but at a minimum covering the driver's reaction time plus some safety buffer.

The key modules incorporated in our architecture and utilized to manage transitions between driving modes are: (1) an operational design domain monitoring system, (2) environment and DSM systems, (3) a trajectory planning system, (4) a system for driving mode assessment and finally (5) a module for DIPA (details can be found in (Insua et al., 2021)). The core system periodically issues predictive risk assessments based on compliance with the ODD. If the predictive probability of exceeding the ODD is sufficiently large, the ADS alerts the driver, and assesses the automated and manual driving modes. If automated is preferable, as the ADS is critically approaching its design limits, the system should enter the emergency mode and issue the appropriate alert. If manual mode is preferred, an RtI is issued to the driver through the HMI followed by a DIPA. If the driver performs too poorly, as assessed from the DIPA model, the driver is perceived not to be in good condition and the emergency mode is triggered. Otherwise, the driver takes back control until further notice.

Algorithm 1 summarises the complete ADS management procedure. There are three modes AUTO, MANUAL, EMERG. Commands on the same line are processed in parallel. The variables ψ_0 and ψ_1 designate the assessment (k steps ahead) of the AUTON and MANUAL modes, respectively.

Algorithm 1: ADS controller.

Input: Priors for ODD, environment, driver state. Utility function
Output: Trajectory from ORIGIN to DESTINY (and implementation of commands when in AUTO or EMERG modes).
while DESTINY not reached **do**
  Read internal sensors. Read external sensors.
  Forecast Environment k steps ahead. Forecast driver
..... state k steps ahead. Compute trajectory.
  Assess driving modes (ψ_0, AUTO; ψ_1, MANUAL).
..... Issue WARNINGS.
  Manage from DRIVING MODE. If DIPA pending,
.....resolve
end while

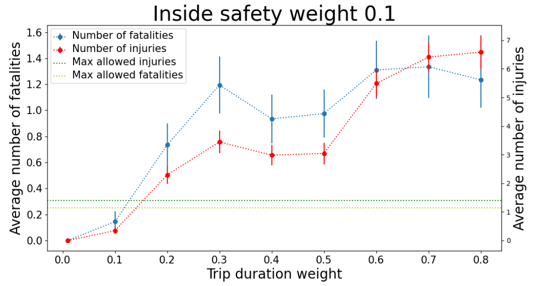
Our driving mode management framework is rooted in statistical decision theory, (French and Insua, 2000). Therefore, the selected utility function to assess driving modes and inform ADS' decisions is critical. One of the most contentious topics in ADS research relates to their decision making in potentially fatal situations, particularly the ethics associated with their automated decision making. As is generally the case with revolutionary technologies, the widespread adoption of ADSs is accompanied by numerous moral uncertainties. Unfortunately, in addressing these dilemmas, decision makers are forced to grapple with unenviable ethical quandaries. Notably, (Awad et al., 2018) developed an online experimental platform called the *Moral Machine* wherein users repeatedly resolve trolley problems to gain insight into societal ethical preferences. Alternatively, we propose using a generic multi-attribute utility model for ADS management that would allow designers, owners and policymakers to tailor ADS behavior according to their own ethical position. The selected objectives, attributes and structure of the multi-attribute utility determine the ethical perspective adopted and can accommodate multiple ethical viewpoints as in (Keeney, 1984). Managing ADS decisions in this way allows for ethical, operational, and regulatory trade-offs to be developed and studied in a computationally tractable manner. Moreover, this research furthers the collective study of ADS ethics and provides a means, in the far-term, to inform regulation. Conversely, the near-term aim of this research is the provision of decision-analytic support for ADS design and operations. As these cover both short- and long-horizon decisions in a dynamic environment, the same set of objectives are considered in multiple decision making contexts.

The ethics associated with a multi-attribute utility function are primarily determined by (1) the preference model's functional form, (2) the selected objectives, (3) the associated attributes, and (4) the objective weights utilized. The proposed objectives and attributes are presented in Table 1, where we have included natural, constructed and proxy attributes in the standard decision analytic description. The preference model's functional form follows the principles of normative decision theory. As it is natural that the preferences of most stakeholders will satisfy mutual preference and utility independence (Keeney et al., 1993), we propose using multiplicative multi-attribute utility functions, see (Caballero et al., 2021a).

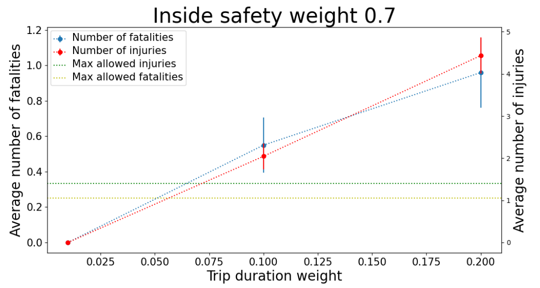
An interesting fact of managing ADS decisions under a normative decision theoretic approach is that the effect of objective weights on liability can be readily observed, especially if weights are tailorable by different stakeholders. Consider the following. Should we assume that a normative framework is utilized to model ADS preferences, then it is highly likely that a government regulator would desire to limit the range of objective weights specified in ADS operations. Responding to this regulation, an ADS manufacturer is likely to sell its vehicles with recommended baseline weight settings, and it is conceivable that the ADS operator would be empowered to tailor these to their needs. However, by adapting the weights, the operator may incur more liability, especially if the ADS's algorithms are modified beyond their legal limits. Indeed, a main advantage of the proposed multi-attribute framework is that it sheds transparency on the decision making process taken during the design of ADSs. Utilizing this proposal, regulators can undertake in-depth simulations of various configurations until they arrive at socially acceptable results. Such configurations can be mandated by law or recommended as industry standards.

As an example, we illustrate how this framework can be leveraged for liability concerns. Assume a regulator has set some safety criteria that must be met by any ADS system to operate on public infrastructures. The safety criteria does not distinguish between individuals inside and outside the vehicle (i.e., both are equally weighted) and is expressed e.g. as follows: *Mean plus two standard deviations of number of injuries and fatalities per X kms should not be greater than 1.4 and 0.25, respectively*. From this criteria, the regulator wishes to determine a recommended industry standard for the ADS objective weights.

Simulating ADS operations, different values for the weights of each component of the multi-attribute utility function can be analyzed, and the regulator can determine whether they met the criteria. Suppose the



(a) Inside safety weight of 0.1



(b) Inside safety weight of 0.7

Figure 3: Average number of injuries and fatalities vs. trip duration weight.

chosen objectives from Table 1 are trip duration and inside and outside safety. Figure 3a shows that if the inside safety weight is fixed at 0.1, trip duration weights greater than or equal to 0.2 do not meet the regulator's criteria. Trip duration weights below 0.2 could be further explored in order to identify the maximum weight fulfilling the safety constraints. In this particular case, a trip duration weight of 0.1 is admissible.

Having completed their analysis, the regulator sets this weight combination as a standard (0.1, 0.1, and 0.8). However, an auto manufacturer determines it can gain market share by allocating more weight to inside safety while maintaining the trip duration weight constant, thereby decreasing the outside safety weight. Concretely, suppose that an inside safety weight of 0.7 is selected. In this scenario, if an injury or fatality occurs, a natural question would be whether or not the auto manufacturer is liable. To make this determination, a simulation of the particular weight configuration could be undertaken to ascertain whether the results fall within the prescribed safety bounds. Figure 3b shows that, in this particular example, the selected weights do not meet the regulator's criteria and thus the manufacturer could reasonably be deemed liable.

As mentioned, the RtI would be communicated to the driver through an HMI. Multiple authors have examined the effect of HMIs on RtIs e.g. (Walch et al., 2015), (Eriksson and Stanton, 2017). We sketch now

Table 1: General objectives and attributes for ADS management.

Objective	Natural attribute	Constructed attribute	Proxy attribute
Min. fuel consumption	Monetary units		
Min. trip duration	Temporal units Monetary units		
Min. driver/passenger discomfort		Yes	ADS movement
Min. injuries of individuals inside (outside) ADS	Number of injuries	Yes	No. in hospital
Min. fatalities of individuals inside (outside) ADS	Number of fatalities VSL	Yes	
Max. respect for inside (outside) ADS	Probability of death/injury VSI	Yes	
Min. damage to ADS	Monetary units		
Min. infrastructure damage	Monetary units		
Min. environmental impact (global/local)	Monetary units Emissions		
Min. harm to manufacturer reputation		Yes	Media salience
Min. harm to societal perceptions		Yes	Media salience

how we assess HMIs from the Trustonomy perspective.

4 HUMAN MACHINE INTERFACES

HMIs are fundamental components of an automated vehicle design and the main channels of information between the vehicle and the driver. Since all automation levels between level -1 and -4 require at least seldom driver interaction, the development of advanced HMI designs is pivotal for establishing a smooth interaction between the driver and the ADS. In particular, this is essential in RtI scenarios. The notifications and information conveyed to the user should be clear and understandable, and indicative of the occurring driving situation. Thus, high quality HMIs capable of generating the appropriate information are of great importance in automated driving.

Road safety issues and the limited amount of information conveyed to the driver directed the focus of automotive industry in developing human-centered HMI design and assessment approaches. Many guidelines regarding HMI quality and assessment criteria have been proposed e.g., (Naujoks et al., 2019), (Carsten and Martens, 2019). Three fundamental features are usability, distraction and acceptance (François et al., 2017). Towards this direction, the vision of the Trustonomy project is to develop an HMI assessment framework considering driver performance, trust and acceptance.

The proposed assessment framework consists of a time-based assessment approach, as well as an innovative *in silico* ergonomics evaluation. Validation of the framework compels evaluating HMIs in suitable automotive environments (real ADS or simulators) by a representative population sample with diverse characteristics (e.g., age, gender, driving experience) to achieve statistical significance. Driver and vehicle data recorded during interactions between the participants and the available HMI designs in the project pilots will be used as input for both framework modules. The time-based module aims to evaluate the impact of HMIs on driver performance and the proposed methodology includes a collection of subjective and objective measurements. Each trial starts with an adaptation driving scenario and, at the end of it, the participant completes a simulation sickness questionnaire. Afterwards, the actual driving scenario initiates including secondary tasks and RtIs, signaled by different HMI modalities. During this time period objective measures are collected such as the time response to the HMI stimuli. At the end of the driving scenario, a subjective assessment follows materialized through comprehension (questionnaire designed based on ISO9186), usability (tailored questionnaires) and workload (modified NASA-TLX) tests. Moreover, the influence of secondary visual-manual tasks will be assessed using the surrogate reference task (SURT).

On the other hand, the ergonomics module of the HMI assessment framework undertakes a more technical approach utilizing software tools to investigate the impact of the driver-HMI interactions on the driver's musculoskeletal system. A biomechanical analysis is conducted using the OpenSim software and appropriate musculoskeletal models that are scaled to capture the anthropometric data of different population classes. The pipeline requires motion capture (MoCap) data that are recorded during the interactions between the driver and the HMI design. The ergonomics method distinguishes between static and dynamic posture analysis. The former aims to assess the installation position of distinct HMI design elements and the level of discomfort imposed to the driver by estimating standard ergonomic indices such as RULA, LUBA and NERPA. The latter evaluates the entire set of interactions through the dynamic evolution of ergonomics indices such as joint energy and power, mean joint torque, angular impulse, and range of motion (Kaklanis et al., 2013) (Risvas et al., 2020). The results are displayed through heatmaps and detailed plots and provide means of comparison between different HMI designs.

The two proposed approaches work separately, yet in a supplementary way, to provide assessment of the various physical and cognitive factors that affect the interactions between the driver and HMI design in an automotive environment. The overall proposed framework can be utilized by vehicle manufacturers to improve their HMI designs. Moreover, the proposed tools can be applied in simulator ADS environments, thus assisting in development of HMI prototypes with human-centered high quality standards.

5 TRUST AND ACCEPTANCE

The concept of trust is fundamental to many of our everyday interactions, and is especially critical when there is uncertainty or incomplete information (Swan and Nolan, 1985). The responses that humans demonstrate towards technology are quite similar in some respects to responses to other humans (Reeves and Nass, 1996); we possibly use habits we have already formed in interpersonal situations. Trust has been identified as a key factor influencing acceptance and reliance on ADS, and in particular in determining the willingness of a human operator to rely on automation in situations of uncertainty (Lee and See, 2004). The full safety and economic potential of ADSs will not be reached if drivers do not accept or use them in an appropriate way. Hence, understanding what factors associate with trust in ADSs is important in understanding and improving human interactions with them. An individual's personality, previous experience and the context in which they are driving are factors that certainly influence trust in ADSs (Hoff and Bashir, 2015). In addition, perceived ease of use, usefulness, safety and privacy risks also mediate trust (Zhang et al., 2019).

Within Trustonomy, and as a precursor to a driving simulator study, an online survey was carried out to explore the "mood music" regarding trust in ADSs. The general public may have heard media reports of ADSs being involved in safety-critical incidents. Do drivers have varying amounts of trust in ADSs as expressed by proxy measures such as willingness to remove their hands and feet from the controls, as well as engage in tasks which direct their gaze away from the road scene. And are there any general conclusions that can be drawn with regards the residents of different countries – are some more inclined to trust ADSs in various situations than others?

Approximately 800 participants (drivers) were recruited to take part in an online survey in the UK, France, Italy and Poland. As well as ascertain demographics, attitudes to automation and personality, respondents rated their trust in ADSs in eleven road environments which varied by road type (urban/rural/motorway), complexity (links, curves or intersections) and road users (present or absent). Respondents were asked to "Imagine you are in the driving seat of an ADS and the ADS is in control. Below are a number of driving scenarios and for each, we would like to know how likely you would be to take your hands off the steering wheel, rest your feet away from the pedals, or take your eyes off the road ahead for longer than you normally do e.g. a couple of seconds or more." An 11-point Likert-type scale from 0=Not at All Likely to 10=Extremely Likely was presented and levels of trust were compared across gender, age, country and road scenario, see Figure 4.

Overall, respondents reported that they would remove their hands and feet from the controls of the vehicle more readily than take their eyes off the road, with males being more trusting than females. UK citizens were least likely to disengage and Italians the most likely. France and Poland were similar to each other. As might be expected, drivers were less likely to disengage on urban roads, compared to rural and motorways and also when the road was complex in design or had other road users present.

With the caveat of this being a self-report survey, we tentatively suggest there may be between-country

Figure 4: Sample of environments presented in the survey.

differences with regards trust, as gauged by proxy measures of disengagement. In addition, there may be (currently justified) reticence in trusting ADSs in conditions where a human operator is perceived to be performing an important monitoring and decision-making role, such as in complex road layouts or where other road users are present.

6 DISCUSSION

We have introduced some of the concepts that, from an interdisciplinary perspective, are being developed within the Trustonomy project to increase acceptance of ADS. A framework to support ethical decisions concerning RtIs was sketched; the entailed decisions need to be properly communicated through appropriate HMIs, demanding adequate assessment; these, in turn, should increase trust and acceptance in ADS, favouring their adoption and their entailed benefits.

ACKNOWLEDGEMENTS

Work performed in the context of the Trustonomy project, which has received funding from the European Community's Horizon 2020 research and innovation programme under grant agreement No 815003.

REFERENCES

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, Jean-François, R., and Iyad (2018). The moral machine experiment. *Nature*, 563(7729):59–64.

Burns, L. and Shulgan, C. (2019). *Autonomy: The Quest to Build the Driverless Car—And How It Will Reshape Our World*. ECCO.

Caballero, W. N., Insua, D. R., and Naveiro, R. (2021a). Ethical and operational preferences automated driving systems. *Tech. Rep.*

Caballero, W. N., Rios Insua, D., and Banks, D. (2021b). Decision support issues in automated driving systems. *Int. Trans. Oper. Res.*

Carsten, O. and Martens, M. (2019). How can humans understand their automated cars? hmi principles, problems and solutions. *Cognition, Technology & Work*, 21.

Eriksson, A. and Stanton, N. A. (2017). Takeover time in highly automated vehicles: noncritical transitions to and from manual control. *Human factors*, 59(4):689–705.

François, M., Osiurak, F., Fort, A., Crave, P., and Navarro, J. (2017). Automotive hmi design and participatory user involvement: review and perspectives. *Ergonomics*, 60(4):541–552. PMID: 27167154.

French, S. and Insua, D. R. (2000). *Statistical Decision Theory*. Edward Arnold.

Hoff, K. A. and Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3):407–434.

Insua, D. R., Caballero, W. N., and Naveiro, R. (2021). Managing driving modes in automated driving systems. *arXiv preprint arXiv:2107.00280*.

Kaklanis, N., Moschonas, P., Moustakas, K., and Tzovaras, D. (2013). Virtual user models for the elderly and disabled for automatic simulated accessibility and ergonomy evaluation of designs. *Universal Access in the Information Society*, 12.

Keeney, R. L. (1984). Ethics, decision analysis, and public risk. *Risk Analysis*, 4(2):117–129.

Keeney, R. L., Raiffa, H., and Meyer, R. F. (1993). *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press.

Lee, J. and See, K. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50–80.

Mahmassani, H. S. (2016). 50th anniversary invited article—autonomous vehicles and connected vehicle systems: Flow and operations considerations. *Transportation Science*, 50(4):1140–1162.

Naujoks, F., Wiedemann, K., Schömig, N., Hergeth, S., and Keinath, A. (2019). Towards guidelines and verification methods for automated vehicle hmis. *Transportation Research Part F: Traffic Psychology and Behaviour*, 60:121–136.

Reeves, B. and Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people*. Cambridge University Press.

Risvas, K., Pavlou, M., Zacharaki, E., and Moustakas, K. (2020). Biophysics-based simulation of virtual human model interactions in 3d virtual scenes. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 119–124.

Swan, J. E. and Nolan, J. J. (1985). Gaining customer trust: a conceptual guide for the salesperson. *Journal of Personal Selling Sales Management*, 5(2):39–48.

Walch, M., Lange, K., Baumann, M., and Weber, M. (2015). Autonomous driving: investigating the feasibility of car-driver handover assistance. In *Proceedings of the 7th international conference on automotive user interfaces and interactive vehicular applications*, pages 11–18.

West, M. and Harrison, J. (2006). *Bayesian forecasting and Dynamic Models*. Springer.

Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R., and Zhang, W. (2019). The roles of initial trust and perceived risk in public's acceptance of automated vehicles. *Transportation Research Part C*, 98:207–220.