Trust Dynamics in Human Autonomous Vehicle Interaction: A Review of Trust Models

Chandrayee Basu and Mukesh Singhal
Cloud Lab, Department of Electrical and Computer Engineering
University of California at Merced
Merced, California 95343

Abstract

Several ongoing research projects in Human autonomous car interactions are addressing the problem of safe co-existence for human and robot drivers on road. Automation in cars can vary across a continuum of levels at which it can replace manual tasks. Social relationships like anthropomorphic behavior of owners towards their cars is also expected to vary according to this spectrum of autonomous decision making capacity. Some researchers have proposed a joint cognitive model of a human-car collaboration that can make the best of the respective strengths of humans and machines. For a successful collaboration, it is important that the members of this human-car team develop, maintain and update each others behavioral models. We consider mutual trust as an integral part of these models. In this paper, we present a review of the quantitative models of trust in automation. We found that only a few models of humans' trust on automation exist in literature that account for the dynamic nature of trust and may be leveraged in human car interaction. However, these models do not support mutual trust. Our review suggests that there is significant scope for future research in the domain of mutual trust modeling for human car interaction, especially, when considered over the lifetime of the vehicle. Hardware and computational framework (for sensing, data aggregation, processing and modeling) must be developed to support these adaptive models over the operational phase of autonomous vehicles. In order to further research in mutual human - automation trust, we propose a framework for integrating Mutual Trust computation into standard Human - Robot Interaction research platforms. This framework includes User trust and Agent trust, the two fundamental components of Mutual trust. It allows us to harness multi-modal sensor data from the car as well as from the user's wearable or handheld device. The proposed framework provides access to prior trust aggregate and other cars' experience data from the Cloud and to feature primitives like gaze, facial expression, etc. from a standard low-cost Human - Robot Interaction platform.

Introduction

Interaction design for safe human - autonomous vehicle co-existence on road must address several scenarios of interactions which include:

1. Semi-autonomous car learning driver intention, which we call in-car interaction.
2. Drivers on semi-autonomous cars (assuming most of the modern cars have alerting capabilities like lane departure, automatic cruise control, etc.) reading the intent of surrounding autonomous cars, an out-of-car interaction.
3. Semi-autonomous car - autonomous car interaction, another out-of-car interaction.
4. Autonomous car reading other car drivers intent.
5. In semi-autonomous cars, which flip between full autonomy and manual, drivers should also understand car intent for safety reasons.
6. Neighboring autonomous vehicles transmitting its context information, like pedestrian obstacle, to another autonomous vehicle.
7. The last but not the least is the two-way interaction between manual and autonomous vehicle.

Automation in car can vary across a continuum of levels at which it can fully or partially replace manual tasks in driving. These tasks include steering control, adaptive cruise control and collision management. Parashuraman *et al.* represented these interactions on a 10-point scale and applied to automation at each stage of human information processing, viz., acquisition, analysis, decision and action (Parasuraman, Sheridan, and Wickens 2000). At the highest level of this scale, a machine has the absolute capability of decision-making and does not solicit human support. At the 5^{th} point on this scale, the machine provides a suggestion and executes it only if the human approves. This model of the levels of automation does not accommodate switching roles of analysis and decision making between human and the Artificial Intelligence (AI) agent as autonomous car.

Some authors, on the other hand, proposed a joint cognitive model for control sharing between autonomous vehicles and its users. They identified the domains of collaboration that can leverage the known abilities of machines and humans (Miller and Ju 2015). For example, human abilities include *reading intent of other road users through eye contact (say), resolving ambiguous situations, resolving novel situations, making ethical decisions*, whereas the machine abilities are *maintaining vigilance, reacting quickly to known situations, sensing in poor visibility or through soft obstacles, controlling the vehicle at limits of traction,*

all aspect sensing and situational awareness. Machines can be designed to pull a wide array of information about a given situation, beyond the limits of experience of a single human, to react quickly to known situations and to be awake/vigilant (most times). A collaboration model that makes the best of the abilities of the team members allows each member to intervene in suitable contexts. For example, a human driver may intervene when he/ she sees a certain kind of terrain because the cars cruise control has failed to meet his expectation once before on a similar terrain. Likewise, a car based on contexts and situational parameters like timeliness of reaction, can decide whether to involve human driver in a decision making.

The paradigm of a collaborative system of human and car naturally brings us to the necessity of maintaining a mutual model of car and human, as pointed out by some prior researchers (Parasuraman, Sheridan, and Wickens 2000),(Sheridan 1997), (Argall and Murphey 2014), (Miller and Ju 2015). While each individual in this collaboration has the power to intervene in certain situations, the goal of a successful collaboration would be to minimize intervention through development of trust. We, therefore, consider trust as an integral part of mutual behavioral models. In this paper we present a review of the existing models of trust in human automation interaction and discuss the relevance of these models in the context of human autonomous vehicle interaction. Most of these models account for the evolving dynamics of human trust. However, these models do not support mutual trust where a machine also maintains its model of trust in its human partner. Moreover, these models are not online yet. We conclude that there is a need for a robust computational framework including sensing, data storage and processing that can support the adaptive models of mutual trust over the operational phase of the autonomous vehicle.

In the following sections we address the concept of mutual trust in team work, and how it can be applied to improve human car collaboration over time. We review existing quantitative models of trust in human automation domain and attempt to establish the importance of developing hardware and computational framework that support monitoring and evolution of user - car mutual trust over time.

Team models of human - automation

Modeling in human car interaction, like in semi-autonomous vehicles, have largely been *one-directional*, like modeling of drivers intent. A team in Berkeley was able to predict human intent of lane changing with 92 % accuracy based on driving simulation studies with real humans (Rutkin 2015). Within the framework of another project, called Brain4Cars (Jain 2015 2015), Jain *et al.* applied Deep Learning on several sensors to infer drivers intent of changing lane and making turns. The goal of the work was to anticipate driver maneuvers in a timely manner for an alerting system. Using videos of the driver inside the car and the road in front, the vehicles dynamics, global position coordinates (GPS), and street maps from a diverse data set with 1180 miles of natural freeway and city driving, the authors showed that their Autoregressive Input-Output HMM model could anticipate maneuvers 3.5 seconds before they occur with over 80 % F1-score in real-time. The authors in a more recent study have used Recurrent Neural Network based sensor fusion for predicting driver maneuver (Jain et al. 2015). This architecture combines sensor - fusion with short term future prediction and achieved a precision of 90 % and a recall of 84 %. In a tangential study, Kim textitet al. used sensor and human-annotated data from 15 drivers, including vehicle motion, traffic states, physiological responses and driver motion to estimate when to interrupt drivers with new information with 94 % accuracy (Kim, Chun, and Dey 2015).

Nikolaidis *et al.* developed a set of generic human type models in the context of larger multi-human and robot team collaboration (Nikolaidis et al. 2015). Based on watching human teams execute a given task, user data was clustered and human type was modeled in terms of individual reward functions learned using inverse reinforcement learning. This is an online learning approach where the initial parameters are uncertain and can be improved through interaction.

As indicated earlier in this section, most of the modeling work in human automation interaction has focused on the driver. Success of human teamwork, however, depends on mutual modeling, a reciprocal ability to establish mental model of the other. Mutual modeling refers to models like *What does he know about what I know?* Trust also falls within the same domain of social behavior. It embodies predictability of the other persons performance in a certain situation and associated belief. While such a process is innate for humans, machines like cars or any AI agent must learn this model. One of the first works that included a comprehensive discussion on mutual modeling in the context of robotics, explored mutual modeling from three academic disciplines, viz., Developmental Psychology, Psycholinguistics and Collaborative Learning and offered relevant sets of experiments and modeling paradigms for robotics (Lemaignan and Dillenbourg 2015). In the next subsection we present some of the existing models of trust in human-automation interaction. We find that, while these models allow for short term trust dynamics, they do not consider the domain of mutual trust.

Trust Models

Trust is considered to be of paramount importance in social acceptability of autonomous vehicles. However, very few experiments in human autonomous vehicles interaction research have been designed to accommodate and infer trust over an extended period of time. Johns *et al.* developed an interface for transfer of control between car and human for steering control and speed control, in which they simulated several modes of danger that would require manual take over and human awareness (Johns, Sibi, and Ju 2014). The allowed transition time for control shift was 7 seconds. The authors did not find significant difference in the driver

awareness between different modes of control. It may be difficult in a short term test to see this difference because as manual drivers it is natural to be aware all the time until humans can develop full trust in the automation over time. For technical as well as anthropomorphic reasons, trust is a dynamic process.

Researchers found that anthropomorphic reactions of humans to robots evolve over three phases, Initialization, familiarization and stabilization (Lemaignan, Fink, and Dillenbourg 2014). In the pre-interaction phase, humans build an *initial capital of anthropomorphism.* During interaction, the level of anthropomorphism increased due to novelty effects and then drops to a stable. While the authors have not mentioned about the exact time scale of this interaction, experimental details suggest at least a few days. The length of the time period over which trust changes will depend on the exact robotic task or a complex set of tasks and the potential of encountering novel situations in the interaction phase. Therefore, we need hardware and computational systems that support large scale experiments on evolving trust over time, particularly as first drivers of autonomous vehicles would have just transitioned from manual or semi-autonomous driving.

Very few models of any trust in human autonomous vehicle context exist, let alone complete models of mutual trust. We found three quantitative models of humans trust in automation, most of which corroborated that trust is a dynamic process, even on a short time scale of interaction.

Lee and Moray developed a dynamic model of human trust in automation based on Auto-regressive Moving Average (Lee and Moray 1992). For 3 days 2 hours per day sessions of running an orange pasteurizer automated plant, the authors recorded the trust dynamics of several operators. Trust was tested by inducing faults in the plant operations. Trust was measured using Muirs 10-point scale reporting (Muir 1987). The authors found that success of collaboration (measured as efficiency) improved constantly as the operators became familiar with the plant operation and loss of trust in case of transient faults were proportional to the magnitude of the faults, but trust recovery was less affected by this magnitude except in case of severe faults.

Some researchers attempted to understand the impact of timing of reliability drop on real-time human trust in machine when human was the sole machine operator. They also tested how disclosing the confidence of the robot of its own sensors affected the dynamics of trust and whether the type of feed- back on this confidence value mattered to the human (Desai et al. 2013). The study found that real-time trust cannot be reflected by traditional trust questionnaires such as Muir questionnaire. Furthermore, trust is affected more by early on robot failure than by later losses of reliability. They developed a trust curve over time that fluctuated based on the performance of the robot and trust was measured as *Area Under Trust Curve*.

Xu *et al.* modeled human trust in automation as a continuous interval representation from complete distrust to absolute trust (Xu and Dudek 2015). The goal was to develop *adaptive trust seeking robots* that can predict human trust at a given situation and adapt its actions and social behavior to improve the trust. Using results of simulation study with humans flying aerial vehicles, the authors developed *OPTIMo*, a dynamic Bayesian model of trust as a function of vehicles performance, frequency of human interventions, previous state of trust and self-reported trust states as trust gained, trust lost or trust unchanged. Their experimental findings suggested that trust varied over time through more interactions and trust was more user-dependent than dependent on actual robot failure. This means, that it is important to develop a *personalized model of human - robot trust* that can accommodate the *temporal dynamics* of this social relationship. A reverse model of trust was proposed by (Argall and Murphey 2014), where the automation infers its own degree of trust in human instructions in the context of mutually controlled automation systems. The goal here was to be able to cede appropriate amount of control to a human instructor based on inferred trust level. In this approach the authors first simulated the initial control behavior via optimal control of a physical AI agent. Human instruction is used in the next step to obtain physical guidance for corrective demonstrations. The resultant stability state of the system following human corrective instructions is then verified. This verification serves as an estimate of trust in the teacher's instructions.

Trust in most of these studies have been measured through human interventions or real-time interruptions as feedback request. Such methods of trust data acquisition may not be sustainable. In order to facilitate modeling and monitoring of mutual trust over the operational phase of the autonomous vehicle new sensing and computational infrastructure must be developed. Several in car sensing is being proposed by researchers for identifying the state of the driver. These sensors can also be leveraged in inferring trust. In the following subsection, we talk about some of the potential sensing systems.

Sensing for trust modeling

Trust has been measured based on self-reporting during or after humanrobot interaction or passively from interactions, say, using frequency of interventions. All of the above approaches to trust inference are *reactive*, in that trust is measured in response to the result of an action or an event, as can be seen in the works of Xu *et al.* and Argall and Murphy. However, a human or a vehicle user may not necessarily have to take any action to modify the state of the autonomous system, in order to infer trust.

Instead of computing trust *reactive* models from user feedback or user intervention, non-verbal cues can be used for *predictive* modeling of trust. In (Lee et al. 2013), the authors used non-verbal social cues like gestures to compute trust between human and an AI agent. The authors conducted a set of game based experiments using Give-Some Game, first between two human partners and then between a human and an AI partner. Give-Some Game is like Prisoner's Dilemma where the players exchange money or tokens. In this work, the number of tokens given by a person

to his or her partner was used as a ground truth trust data. Initially videos were manually coded for several non-verbal cues like leaning forward or backward, touching different parts of body, head and face, head signals like shaking, nodding, smile signals, eye contact and arm signals. Using a Support Vector Machine (SVM) model the researchers concluded that *joint cues* like face touching, arms crossed, leaning backwards and hand touching were most informative of the distrust state (or negative trust) of the human in the AI agent or other human partner. The *duration of occurrences of these joint signals* as well as the *temporal relationship between non-verbal cues*, (for example, smiling $\rightarrow$ face touching $\rightarrow$ hand touching ...) were found to be strongly correlated with trust. The researchers captured temporal relationships with two different Hidden Markov Models, one for low trust and the second Hidden Markov Model for high trust. They compared the performance of this trust model with features engineered from domain knowledge against that with features learnt from variable ranking. The algorithm using domain knowledge outperformed human prediction as well as that using learnt feature. Later on, Lee *et al.* proposed a software framework to replace manual video coding with automatic detection of gesture primitives like eye gaze, arms position and so on using 3D motion capture(Lee, Knox, and Breazeal 2013). Autonomous vehicles will have a host of sensors that can be leveraged to record such non-verbal cues of human trust as the above study. However, the most informative non-verbal cues of trust in the context of autonomous vehicles may be different from those in the closed lab environment, where the interaction is between two players of economic games. It should also be possible to interpret human trust in an autonomous vehicle agent from the emotional state of the human.

The same built-in sensor suite in the car may be deployed to infer emotional state of the user. Jain *et al.* could predict driver intent 3.5 seconds before maneuvers using in car video camera. It may also be relevant to infer a driver or a user's state of trust using physiological sensors. For example, on seeing a novel kind of obstacle or when a car's behavior does not match the users mental model of car's performance, an alert driver may show physiological signs of distrust in the form of stress. In order to estimate drivers emotional states, researchers are exploring several sensing options (Kim, Chun, and Dey 2015), (Ji, Zhu, and Lan 2004), which include head and gaze tracking for attention (Smith, Shah, and da Vitoria Lobo 2003), pupillometry for arousal (Palinko et al. 2010), variations in heart rate (Healey and Picard 2000), stress measurement (Healey and Picard 2005), force sensors built into the seats for postures and visual sensing. As mentioned earlier in this paper, besides inferring how much trust a user has in his autonomous vehicle partner, it is also important for the vehicle to understand how much it can trust the instructions of its human partner. Trust, thus defined on a sliding scale can be the deciding factor for ceding control to a human operator or even asking for help from the user in unknown circumstances. This trust will be a function of the level of alertness and physiological states of the user and the mechanical stress in the car resulting from human intervention. Of these, the former that can again be computed from the same set of physiological and camera sensors. These sensors, however, need not be built into the car. The users mobile phone or fitness sensors, for example, can share the raw data and/ or the inference with the car if needed. With some initial days of training with self-reported trust, the above sensing and inference mechanisms can provide longer term trust monitoring without interruptions. Additional sensors will be required to monitor the state of the car like mechanical stress or wear and tear as a result of human action or intervention. To allow further research in mutual human-autonomous vehicle trust and to validate existing trust models, a framework that integrates multi-modal sensing, data aggregation and interface design is of utmost importance. In the following section, we propose such a framework.

Framework for Trust Modeling

We propose a potential framework for mutual trust that can enable further research on trust between a human and his/her autonomous vehicle partner and integration of a trust module in human - car interaction experiments. The framework should be flexible enough to accommodate diverse sensing modalities and *active, passive, reactive or predictive* inference of trust between a human and autonomous vehicle. Furthermore it should provide adequate storage to maintain and update *personalized* model of *dynamics of trust* over the life time of the vehicle. This framework encompasses two different trust modules, for user's trust and vehicle's trust (which we call agent trust henceforth) respectively.

User Trust - The Autonomous vehicle deploys this software module to infer how much trust the user has in the vehicle performance, at real time, and aggregated over a time frame using frequency of interventions, non-verbal cues like gesture primitives and physiological primitives, which are learnt in turn from in-car and user-mounted sensor data. Trust can be computed on a sliding scale as proposed by (Desai et al. 2013).

Agent Trust - The Autonomous vehicle deploys this software module to infer how much it can trust the user in help solicitation under known or unknown circumstances or cede control to the user. This trust is also computed real time and aggregated over a time window. Trust can be computed based on mechanical stress and stability state of the vehicle in response to following a human decision or predicted based on user's physiological primitives. A Cloud connected vehicle may also be able to compare the instructions or judgement of the user against that of a larger population of other cloud connected agents in similar contexts.

A schematic of the potential components of the Mutual Trust framework in Figure 1. We assume that the sensor suite, the gestural primitives and the physiological primitives can be accessed from a standard Human - Robot Interaction (HRI) platform. The arrows represent data and information flow between the cloud, the HRI platform, car server and Mutual Trust module. Besides real time and aggregated trust, the two other necessary components of the system are interface to access results of the mutual trust models and

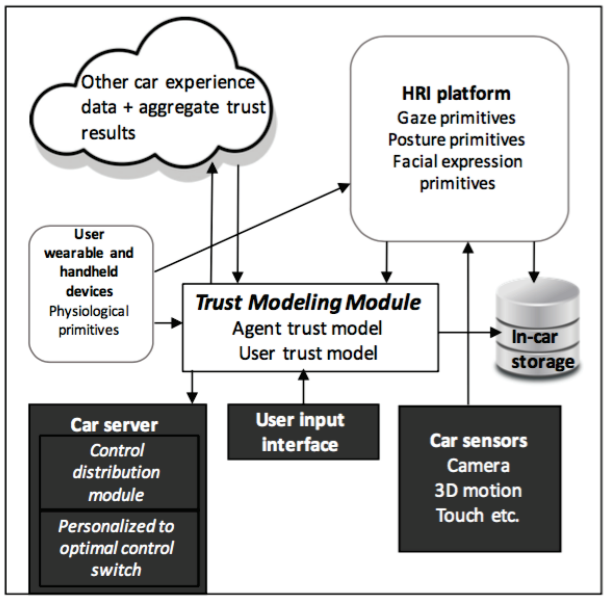

Figure 1: Trust Modeling Framework showing components and information flow

interface to enable human inputs as ground truth in trust research. The results of the mutual trust models can be requested by a *control distribution module* which decides the flow of control between human and autonomous agent or by a *personalized to optimal control switch*. The latter is a proposed module that alters the car control parameters between user preferred and optimal settings. For example, as a user gains trust in an autonomous car, the car may slowly shift from user preferred braking distance to optimal braking distance. The results of the trust model may also be utilized by human - robot interaction interface to generate appropriate communication between the partners during research or operational phase of the car. For example, the *human help seeking* module of the vehicle may actively request trust results of the agent in the user. Likewise, an input interface to record ground truth human trust can be designed to support one or more input types like button press, gesture, facial expressions previously agreed upon and voice feedback.

Discussion

Several researchers in human-automation interaction have agreed upon the importance of mutual modeling for success of interaction. Trust is an indispensable part of this model. In natural human-human interaction a human automatically mentalizes how much trust another team member has in him. For AI agents like autonomous car, reciprocal model is not natural. The AI agent must be designed to develop, maintain and update this model through the course of interaction with humans. Most of the trust models in the literature have focused on the dynamics of a humans trust on automation. These approaches, however, do not allow online modeling as yet. Most of these models are set in an asymmetric control setting, where the human operator still has absolute capability to intervene in the operation of the automation at will. One of the major shortcomings of these models is that, they are not mutual, in that, a machine does not have a model of human. In other words, the AI agent does not know whether it can trust the human partner under certain situations. In a truly collaborative setting of human car driving, where the decision making shifts between the human and the agent based on individual strengths, a mutual model of trust is indispensable. It would therefore be necessary to further develop sensing hardware and a computational framework that can not only acquire information to infer human trust in the car, but is also necessary to model the cars trust in the user or the driver. Furthermore, this computational framework should accommodate the evolution of trust between human and their car over time.

Therefore, we proposed a framework for potentially integrating Mutual Trust computation into a standard Human - Robot Interaction platform. This framework includes User trust and Agent trust, the two basic components of Mutual trust, that can harness multi-modal sensor data from the car, user's wearables or handheld device, prior trust aggregate and other car experience data from the Cloud and feature primitives like gaze, facial expression etc. from a standard low-cost HRI platform. Real time trust results may be stored in in-car database servers. Aggregated trust over a window must be accessed by the car server for allocating control and decision making between the user and the AI agent, for soliciting help under unknown situations and for selecting the degree of personalization to user preferences.

The Mutual Trust Module may be integrated into autonomous vehicle research, experiments and operations in several ways. Some HRI researchers have integrated their HRI software into Robot Operating Systems (ROS) library as ROS nodes. Android phones and iPads have been successfully used by researchers as user input device, robot face and data visualization platforms. Most of the smartphones phones have a multitude to sensors that can enable data acquisition for computation of gesture primitives as well as some physiological primitives like heart rate. In its most basic implementation, therefore, mutual trust module could be a multi-platform app that can be installed on a car server or a smartphone that users or researchers could plug into the vehicle dashboard.

Lastly, trust is a sensitive topic that can affect minor user comfort in the autonomous vehicle to more safety critical decisions. User discomfort may be incurred, for example, in situations where the agent may not be maintaining the user perceived safe distance to braking. More safety critical decisions must be taken when an autonomous vehicle faces an unknown obstacle. Under such circumstances a corrupt trust model prevents transfer of control to the user. However, in Mutual Trust model the control transfer depends on both agent trust and user trust and hence may be more robust, particularly when the car has access to data from other cars' experiences.

Conclusion

In this paper we made an attempt to establish a case for mutual modeling of trust in human autonomous vehicle inter-

action for a successful driving experience and presented a review of quantitative models of human trust in automation from literature. Assuming autonomous vehicle driving will be a true teamwork that will leverage respective strengths of the human and machine partners, we conclude that current experimental research in human car interaction must be improved to accommodate the dynamic nature mutual trust between a car owner and his car. We also found that there is significant scope for future research in the domain of mutual modeling for human car interaction and hardware and computational frameworks (for sensing, data aggregation, processing and modeling) that can support the same over time. We proposed a framework for integrating Mutual Trust computation with a standard Human - Robot Interaction modeling and research platform. This framework allows to model user trust in car and car's trust in user separately using multi-modal sensor data from the car, user's wearable or handheld device, prior trust aggregate, other cars' experience data from the Cloud and feature primitives like gaze, facial expression etc. from a standard low-cost Human - Robot Interaction platform.

Acknowledgments

The authors are grateful to Postdoctoral scholar Santosh Chandrasekhar for his insights on system design for trust modeling.

References

Argall, B., and Murphey, T. 2014. Computable trust in human instruction. In *AAAI Fall Symposium Series*.

Desai, M.; Kaniarasu, P.; Medvedev, M.; Steinfeld, A.; and Yanco, H. 2013. Impact of robot failures and feedback on real-time trust. In *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on*, 251–258.

Healey, J., and Picard, R. 2000. Smartcar: detecting driver stress. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 4, 218–221 vol.4.

Healey, J., and Picard, R. 2005. Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions on* 6(2):156–166.

2015. Brain4Cars. http://brain4cars.com.

Jain, A.; Singh, A.; Koppula, H. S.; Soh, S.; and Saxena, A. 2015. Recurrent Neural Networks for Driver Activity Anticipation via Sensory-Fusion Architecture. *arXiv preprint arXiv:1509.05016*.

Ji, Q.; Zhu, Z.; and Lan, P. 2004. Real-time nonintrusive monitoring and prediction of driver fatigue. *Vehicular Technology, IEEE Transactions on* 53(4):1052–1068.

Johns, M.; Sibi, S.; and Ju, W. 2014. Effect of cognitive load in autonomous vehicles on driver performance during transfer of control. In *Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '14, 1–4. New York, NY, USA: ACM.

Kim, S.; Chun, J.; and Dey, A. K. 2015. Sensors know when to interrupt you in the car: Detecting driver interruptibility through monitoring of peripheral interactions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 487–496. ACM.

Lee, J., and Moray, N. 1992. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35(10):1243–1270.

Lee, J. J.; Knox, W. B.; Wormwood, J. B.; Breazeal, C.; and DeSteno, D. 2013. Computationally modeling interpersonal trust. *Frontiers in psychology* 4.

Lee, J. J.; Knox, B.; and Breazeal, C. 2013. Modeling the dynamics of nonverbal behavior on interpersonal trust for human-robot interactions. In *AAAI Spring Symposium Series*.

Lemaignan, S., and Dillenbourg, P. 2015. Mutual modelling in robotics: Inspirations for the next steps. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, 303–310. New York, NY, USA: ACM.

Lemaignan, S.; Fink, J.; and Dillenbourg, P. 2014. The dynamics of anthropomorphism in robotics. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, HRI '14, 226–227. New York, NY, USA: ACM.

Miller, D., and Ju, W. 2015. Joint cognition in automated driving: Combining human and machine intelligence to address novel problems. In *AAAI Spring Symposium Series*.

Muir, B. M. 1987. Trust between humans and machines, and the design of decision aids. *Int. J. Man-Mach. Stud.* 27(5-6):527–539.

Nikolaidis, S.; Ramakrishnan, R.; Gu, K.; and Shah, J. 2015. Efficient model learning from joint-action demonstrations for human-robot collaborative tasks. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, 189–196. New York, NY, USA: ACM.

Palinko, O.; Kun, A. L.; Shyrokov, A.; and Heeman, P. 2010. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, 141–144. New York, NY, USA: ACM.

Parasuraman, R.; Sheridan, T.; and Wickens, C. D. 2000. A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 30(3):286–297.

Rutkin, A. 2015. Autonomous Cars are learning our unpredictable driving habits. https://www.newscientist.com/article/mg22730362-900-autonomous\\-cars-are-learning-//our-unpredictable-driving-habits/?utm_source=NSNS&utm_medium=SOC&utm_campaign=twitter&cmpid=SOC\%7CNSNS\%7C2015-GLOBAL-twitter.

Sheridan, T. 1997. Eight ultimate challenges of human-robot communication. In *Robot and Human Communication, 1997. RO-MAN '97. Proceedings., 6th IEEE International Workshop on*, 9–14.

Smith, P.; Shah, M.; and da Vitoria Lobo, N. 2003. Determining driver visual attention with one camera. *Intelligent Transportation Systems, IEEE Transactions on* 4(4):205–218.

Xu, A., and Dudek, G. 2015. Optimo: Online probabilistic trust inference model for asymmetric human-robot collaborations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, 221–228. New York, NY, USA: ACM.