Research Article

Human's Overtrust in and Overreliance on Advanced Driver Assistance Systems: A Theoretical Framework

Toshiyuki Inagaki and Makoto Itoh

Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8573, Japan

Correspondence should be addressed to Makoto Itoh; itoh.makoto.ge@u.tsukuba.ac.jp

This paper gives a theoretical framework to describe, analyze, and evaluate the driver's overtrust in and overreliance on ADAS. Although "overtrust" and "overreliance" are often used as if they are synonyms, this paper differentiates the two notions rigorously. To this end, two aspects, (1) situation diagnostic aspect and (2) action selection aspect, are introduced. The first aspect is to describe overtrust, and it has three axes: (1-1) dimension of trust, (1-2) target object, and (1-3) chances of observation. The second aspect, (2), is to describe overreliance on the ADAS, and it has other three axes: (2-1) type of action selected, (2-2) benefits expected, and (2-3) time allowance for human intervention.

1. Introduction

Driving a car requires a continuous process of perception, cognition, action selection, and action implementation. Various functions are implemented in an advanced driver assistance system (ADAS) to assist a human to drive a car in a dynamic environment. Such functions, sometimes arranged in a multilayered manner, include (a) perception enhancement that helps the driver to perceive the traffic environment around his/her vehicle, (b) arousing attention of the driver to encourage paying attention to potential risks around his/her vehicle, (c) setting off a warning to encourage the driver to take a specific action to avoid an incident or accident, and (d) automatic safety control that is activated when the driver takes no action even after being warned or when the driver's control action seems to be insufficient [1]. The first two functions, (a) and (b), are to help the driver to understand the situation. Understanding of the current situation determines what action needs to be done [2]. Once situation diagnostic decision is made, action selection decision is usually straightforward, as has been suggested by recognition-primed decision making research [3]. However, the driver may sometimes feel difficulty in action selection decision. Function (c) is to help the driver in such a circumstance. Note that any ADAS that uses only

the three functions, (a)–(c), is completely compatible with the *human-centered automation* principle [4] in which the human is assumed to have the final authority over the automation.

Suppose an ADAS contains the fourth function, (d). Then the ADAS may not always be fully compatible with the human-centered automation principle, because the system can implement an action that is not ordered by the driver explicitly. Some automatic safety control functions have been already implemented in the real world. Typical examples are seen in an advanced emergency braking system (AEBS) and a lane departure prevention system (LDP). When a vehicle is approaching to the forward vehicle, AEBS tightens the seat belt and adds a warning to urge the driver to put on the brake. When the system determines that the driver is late in braking, it applies the brake automatically based on its decision. LDP is an automatic system that applies the brakes to individual wheels, without any intervention of the driver, to prevent the vehicle from departing the lane. The fact that the driver may not always be kept as the final authority over the automation in such ADAS does not necessarily mean that those designs should be prohibited. On the contrary, the automatic safety control functions are effective and indispensable to attain driver safety, which suggests the domain dependence of human-centered automation [5]. It is true,

however, that careful investigations are needed regarding to what extent the system may be given authority for deciding and acting autonomously without asking the human driver's approval or consent, because autonomy of smart machines sometimes brings negative effects, such as the out-of-the-loop performance problem, loss of situational awareness, complacency or overtrust, and automation surprises; see, for example, [6–10].

Moreover, as for the fourth function, (d), the following question is frequently asked: "when the ADAS is capable of coping with the situation automatically without any intervention of a driver, is not it possible for the driver to be overly reliant on the system and give up active involvement in driving?" For instance, the Ministry of Land, Infrastructure, and Transport as well as the National Police Agency of the Government of Japan has been somewhat discreet in introducing highly automatic safety control functions into ADAS on concern that the drivers may place "*overtrust*" in or "*overreliance*" on automation. However, discussions regarding overtrust and overreliance have not been rigorous enough yet until this point. As ADAS becomes smarter and more autonomous, these issues attract more serious concerns worldwide, for example, ASV project in Japan, HAVEit, and ISi-PADAS projects in EU.

Aviation domain has various studies regarding overreliance on automation; see, for example, [11–14]. Suppose that the automation is very rare to miss detections (i.e., it almost always alerts the human when an anomaly or an undesirable event occurs). Although a given alert is likely to be false, the human can be confident that there is no undesirable event as long as no alert is given. The human accordingly does not take precautions while the automation gives no alert. Meyer [13] has used the term *reliance* to express such a response of the human. If the human assumed that the automation will always give alerts when an undesired event occurs, that may be overtrust in the automation's capabilities, and the resulting reliance on the automation can be overreliance.

The relevant term, *complacency*, is usually defined as "self-satisfaction especially when accompanied by unawareness of actual dangers or deficiencies" [15]. However, the term is often used in human factors area to express a phenomenon that the human does not monitor the automation. Moray and Inagaki [16] have pointed out that the usage is misleading, because "not monitoring the automation" does not necessarily mean that the human is complacent. An obvious counterexample is that the human is busily occupied with extremely urgent tasks. Therefore, this paper tries to avoid using the term complacency.

This paper proposes a theoretical framework to describe, analyze, and evaluate the driver's overtrust in and overreliance on ADAS. Although the two notions, overtrust and overreliance, are often used as if they are synonyms, this paper differentiates the notions rigorously. To this end, two aspects, (1) situation diagnostic aspect and (2) action selection aspect, are introduced. The first aspect is to describe overtrust, and it has three axes: (1-1) dimension of trust, (1-2) target object, and (1-3) chances of observation. The second aspect, (2), is to describe overreliance on the ADAS, and it distinguishes other three axes: (2-1) type of action selected,

(2-2) benefits expected, and (2-3) time allowance for human intervention.

2. Overtrust

Overtrust can be defined as a psychological state in which the human trust is inappropriately high. Overtrust is an incorrect situation diagnostic decision claiming that the object is trustworthy when it actually is not. This paper introduces three axes for describing the types of overtrust in a precise manner.

2.1. Dimension of Trust. The first axis (1-1) gives the dimension of trust. Lee and Moray [17] have distinguished four dimensions for trust: (a) *foundation*, representing the fundamental assumption of natural and social order, (b) *performance*, resting on the expectation of consistent, stable, and desirable performance or behavior, (c) *process*, depending on an understanding of the underlying qualities or characteristics that govern behavior, and (d) *purpose*, resting on the underlying motives or intents. Trust in an object is appropriate when all the dimensions are evaluated correctly. When there is a dimension that is evaluated inappropriately high, perceived trust is seen as overtrust. Therefore, some types can be distinguished for overtrust depending on which dimension of trust is violated.

Example 1. Suppose the driver thought that "the ADAS has been successful in coping with the situations so far. I am sure that the system will continue to be successful hereafter, too." This is a type of overtrust, violating the second dimension of trust.

Example 2. Imagine a case in which the driver thought that "I do not know how the function is implemented in the ADAS. I am not informed how the task is carried out, either. However, it would be quite alright even if I do not know the details." This is a type of overtrust, violating the third dimension of trust.

Example 3. Assume that the driver said that "I do not understand why the system is doing such a thing. However, the system should be doing what it thinks it necessary and appropriate. The system will not harm us." This type of overtrust does not satisfy the fourth dimension of trust.

Itoh [18] developed a model of human trust in automation and discussed the relationship between the three dimensions of trust, that is, purpose, process, and performance. The model takes into account the function of an automated system, limitation of the working conditions for the function, and the reliability of the automation function within the limitation. User's misunderstanding of the function is related to overtrust in terms of the purpose dimension. Expecting successful work of the automation beyond the limitation is a type of overtrust, violating the dimension of process. On the other hand, human's complete trust in automation within the limitation of the prescribed working conditions may not be overtrust if the reliability of the automation is perfect within

the conditions. Itoh [18] also suggested that increase of trust in terms of performance may result in the overtrust in terms of process, and finally in the overtrust in terms of purpose. Such expansion of overtrust was called "ripple effect."

2.2. Target Object to Which Overtrust Is Addressed. The second axis (1-2) describes a target object to which the driver places inappropriately high trust. This paper distinguishes five types of target objects, computer (C), software (S), hard-ware (H), environment (E), and liveware (L) according to the C-SHEL model [19] describing human interactions with other humans, technology, and the environment; see, Figure 1.

Example 4 (overtrust in computer). The adaptive cruise control system (ACC) performs the longitudinal control on behalf of a driver. Suppose the driver thought that "a car just ahead of me on the next lane may be cutting in. The ACC must have already noticed the car and will adjust the control when appropriate." This is overtrust in the ACC (computer), when the car on the next lane is outside the range of the ACC and the driver does not notice that.

Example 5 (overtrust in software). Imagine a case in which a driver thought that "today is the first day for me to use a brand new system. Oh, I forgot to read the manual. There should be no problem even if I pressed the buttons in a wrong sequence. Fool-proof or tamper-proof functions must be implemented in the software."

Example 6 (overtrust in hardware). Assume that a driver thought that "strictly speaking, this is the time for me to bring the car to a periodic inspection. However, I am quite busy right now and I have never experienced hardware troubles in the car. Why do I have to bring my car periodically for an inspection? My car will not fail."

Example 7 (overtrust in environment). Suppose a man is driving his car thinking that "this road is simply straight. Moreover, there is usually little traffic. It is very relaxing to drive on this simple and somewhat boring road." In reality, environment may alter with time.

Example 8 (overtrust in liveware). Suppose a driver is approaching to an intersection with a blind corner and that an ADAS sets off an alert telling the driver that "a car is approaching to the intersection from the right on the crossing road." The driver cannot see the car himself, because the car is just behind the blind corner. The ADAS generated the alert based on the information obtained via vehicle-to-vehicle or vehicle-to-infrastructure communication technology. Suppose the driver thinks that "I do not see any car. If there is a car, the car will surely yield the right of way, because it is me that is on the priority road," which is overtrust in the driver (liveware) of the other car at the blind corner.

Example 9 (overtrust in liveware). Imagine a car equipped with an electronic stability control system (ESC) that improves stability by applying the brakes to individual wheels
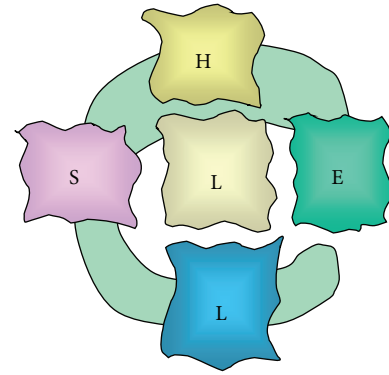
C: computer E: environment
S: software L: liveware
H: hardware

FIGURE 1: C-SHEL model.

when skids or loss of steering control was detected. Suppose the ESC worked at a sharp curve on a slippery road. If the human interface was not properly designed to let the driver know that the ESC was activated, the driver might feel inappropriate confidence on his driving skill, failing to recognize that it was the ESC that assured the stability of the car at the curve. This is a case of overtrust in the driver himself/herself.

2.3. Chances of Observation. The third axis (1-3) distinguishes two classes for ADAS: (a) ADAS for use in normal driving and (b) ADAS for use in emergency. A most prominent characteristic that distinguishes the two classes is the chances to observe ADAS functioning.

Example 10. ADAS for use in normal driving (e.g., ACC) usually aims to reduce the driver workload and works continuously for certain period of time. Since such an ADAS is used daily, the driver observes the system's "intelligent" behaviors repeatedly, which gives the driver a number of opportunities for constructing a mental model of the ADAS.

Example 11. ADAS for use in emergency (e.g., AEBS described in Section 1) usually aims to prevent a catastrophic event from occurring and thus to attain the driver safety. Since such an ADAS is activated only in cases of emergency, it would be very rare for an ordinary driver to see the ADAS works. That suggests that the driver may not be able to accumulate chances sufficient enough for constructing a concrete mental model of the ADAS.

3. Overreliance

Overreliance on an ADAS is a psychological state in which the human reliance on the ADAS is inappropriately high. More precisely, overreliance is an incorrect action selection decision based on an incorrect situation diagnostic decision regarding the ADAS (i.e., the overtrust in it). Here we introduce three axes for describing types of overreliance, that is,

(2-1) type of action selected, (2-2) benefits expected, and (2-3) time allowance for human intervention.

3.1. Type of Action Selected. For the action selection, (2-1), this paper distinguishes the following two types of decisions: commission-like action selection decision and omission-like action selection decision. The former is a selection and implementation of an action that is not suitable to a given situation. Risk compensating behavior [20, 21] could be categorized as a commission-like action. The latter is a failure to select or implement an action that is needed in a given situation.

Example 12 (commission-like action). Suppose a man is driving a car equipped with an ESC at high speeds, which is overreliance on the ESC if it was a clear but extremely cold winter morning and it had rained before dawn. It would be inappropriate to drive a car at high speeds in such an adverse weather condition although the car is equipped with the ESC.

Example 13 (omission-like action). Suppose a man is driving a car by using an ACC and a lane keeping assistance system (LKA). LKA is an automatic system that recognizes the lane and provides the driver with assisting steering torque to keep the car around the center of the lane. Suppose the driver decided to let the LKA take care of the lateral control completely for a while, so that he could consult the navigation system to know how to access his destination. If the LKA was of the type that ceases to control the steering when it determines, through monitoring the driver behavior, that the driver has not been active in steering, the driver's decision to trade the full authority to the LKA is overreliance on the LKA. A case may happen that nobody controls the car, if the human interface did not tell the driver clearly that the LKA returned the authority and responsibility of steering back to the driver based on its decision that the driver had been inactive in steering for a certain period of time.

3.2. Benefits Expected. The second axis is to describe whether the driver can produce some benefits by relying on the assistance system.

Example 14. Suppose the driver assigns the ACC all the tasks for longitudinal control of the vehicle. That may enable the driver to find time to relax muscles and extend legs after stressful maneuvering, or to allocate cognitive resources to find a right way to the destination in a complicated traffic conditions. In this way, relying on the assistance system sometimes brings extra benefit to the driver, when the system is for use in normal driving.

Example 15. AEBS is activated only in emergency, and the time duration for the AEBS to fulfill its function is short, say several seconds. It is thus not feasible for the driver to allocate the time and resources, saved by relying on the AEBS, to something else to produce extra benefit within the several seconds. A similar argument may apply to other assistance systems designed for emergency. If a driver relies on AEBS in normal driving in a sense that the driver lets the AEBS brake when necessary, it would be beneficial for the driver to be able to decrease his/her vigilance and to be relaxed (benefit from an omission-like action). On the other hand, the driver could increase the vehicle speed and reduce time headway and let the AEBS take care of braking. This is another benefit obtained from a commission-like action.

3.3. Time Allowance for Human Intervention. The third axis is to describe whether the driver can intervene into the assistance system's control when the driver determined that the system performance differs from what he or she expected. Note here that this axis could be used to judge whether a driver's reliance is excessive or not. That is, if the time allowance is large enough, the driver has to intervene into control when necessary. However, this axis is not for explaining causes of overreliance.

Example 16. In case of ACC, it may not be hard for the driver to intervene to override the ACC when its performance was not satisfactory. In fact, a driver has to intervene into control when the deceleration of the lead vehicle is larger than what the ACC can manage. If the driver does not apply the brake him/herself, the driver's reliance is regarded as excessive.

Example 17. In case of AEBS, it might be unrealistic to assume that the driver can intervene into control by the AEBS when he or she decided that the AEBS performance was not satisfactory, because the whole process of monitoring and evaluation of AEBS performance as well as decision and implementation of intervention must be done within a few seconds. Thus, driver's failure to override the AEBS when the system is not successful to avoid a crash does not directly mean that the driver's reliance is excessive, if the driver maintained long enough time headway and paid enough attention to the lead vehicle. If, on the other hand, the driver has maintained short time headway and let the AEBS apply the brake, it should be regarded as overreliance.

4. Possibilities of Overtrust and Overreliance

Let us discuss overtrust in and overreliance on ADAS by integrating viewpoints given in Sections 2 and 3.

4.1. Communication-Based Information Provision. Suppose an ADAS has a communication-based function to set off an alert on a car that the driver may not be able to see. There are some objects in which the driver may place overtrust. Example 8 has described one of such cases, where the driver of some other car (a *liveware* in the target-object axis in Section 2.2) needs to be taken into account from a viewpoint of *performance* in the dimension-of-trust axis in Section 2.1.

Consider a case in which a driver is approaching to an intersection that has blind corners but has no traffic lights. The communication-based infrastructure was installed a year ago. The infrastructure can detect cars travelling on the roads crossing each other, and it sends a signal to an onboard ADAS of a car, so that the ADAS can set off an alert to let the driver know an approach or existence of some car(s) on a crossing road. Suppose the driver drives the road daily

(i.e., chance-of-observation axis) and has been satisfied with the performance (i.e., dimension-of-trust axis) of the communication-based alert. The driver now thinks that "I am sure that no car is coming toward me when no alert is given. Why not cross the intersection without deceleration?" In this case, the driver is overlooking the possibility of hardware failure of the infrastructure (i.e., target-object axis). His situation diagnostic decision that "no car must be approaching toward me because no alert is there" is inappropriate (i.e., overtrust). When the communication-based infrastructure was out of service, no alert can be given to the driver. Thus, the action selection decision to "cross the intersection without deceleration" is overreliance on the function of the communication-based alert, when the driver abandons the responsibility to be vigilant.

4.2. Adaptive Cruise Control System. Conventional adaptive cruise control (ACC) systems are not able to control headway in reference to slow or stopped vehicles [22, 23]. According to an interview to owners of vehicles equipped with an ACC system, some part of the owners did not understand this inability [24]. Itoh [18] conducted a driving simulator experiment to observe overtrust in and its resulting overreliance on the ACC. Participants were requested to drive a car by using an ACC that can control the host vehicle to a complete stop when the lead vehicle decelerates and stops. However, the ACC does not recognize stationary body (such as, cars standing still). Participants experienced 69 drives with ACC during the period of four days. At the final trial on the fourth day, participants were given a case in which, after 20 minutes of following the lead vehicle at 100 km/h, the lead vehicle made a lane change and the host vehicle happened to approach to the tail of a traffic jam, where all the vehicles in the jam stood completely still. Participants needed to apply the brake by themselves. One collision and some near collisions into the car at the tail of the jam were observed in the experiment. None of the participants who caused the collision or near collision were drowsy or distracted. Data analyses and investigation of those cases suggested that the participants developed trust in the ACC while experiencing repeatedly the ACC's successful lead vehicle followings to complete stops (i.e., chance-of-observation axis), and that some participants had inappropriate expectations (i.e., dimension-of-trust axes) that the "ACC would control the host vehicle nicely to a vehicle ahead," even for an already standing still vehicle. The participants' failure in applying the brake (i.e., omission-like action) is due to overreliance on the ACC, induced by the overtrust in it.

4.3. Airbag. Strictly speaking, airbags may not be ADAS. However, it is worth mentioning that problems are related to deployment of airbags, because the problems are closely related to the issues of overtrust and/or overreliance.

Since 1990s, passenger cars have been widely equipped with airbags. However, related to the use of airbags, many troubles occurred in Japan, especially at the early stage of the spread. For example, there were cases in which a driver was killed or seriously injured by the deployment of the airbag when the vehicle crashed into something (see, e.g., [25]). In those cases, the drivers had not fastened their seat belt when they had the accident. One possible reason for the nonuse of the seat belt was that the drivers regarded the airbag as an alternative to a seat belt. However, such understanding is inappropriate. An airbag is a "supplemental restraint system" (SRS) which means that the airbag is supplemental to the seat belt. An airbag itself is not enough to support the driver. Regarding an airbag as an alternative to the seat belt can be said as overtrust in an airbag system in terms of "purpose" dimension of trust. Note here that such overtrust could emerge even if the driver does not have chances to observe cases when an airbag works at all beforehand. The process of emerging such excessive trust in an airbag system is different from the one for ACC systems. Driver's reliance on an airbag without being belted is overreliance. It is an omission-like action in a sense that the driver omits fastening his/her seatbelt. For drivers, being unbelted may be worth doing because it could be relaxing for the drivers. This type of overreliance can be detected by monitoring the state of the seatbelt.

On the other hand, some drivers complained about the nonactivation of an airbag when their vehicle crashed into something. For most cases, the reason for the non-activation was not a malfunction of the airbag, but the situation was beyond the system's operative conditions. For example, the airbag at the driving seat may not deploy in a case of offset crashes. It can be said that the complaint is due to driver's overtrust in process in the systems if the driver thinks just "I do not know how the airbag system works, but it will deploy whenever a crash occurs." This is another type of overtrust in an airbag system. Note here that the driver may fasten his/her seatbelt even if the driver has such overtrust in an airbag system. This type of overtrust may not be found until a crash occurs.

The above two examples suggest that it is necessary to identify what type of overtrust/overreliance which is under consideration.

4.4. Advanced Emergency Brake System. Conventional AEBS did not aim to prevent a catastrophe from occurring but to mitigate collision damages. Troubles due to drivers' overreliance on AEBS have not been reported from field operations in the real world. On the other hand, development of technologies increased the possibility of AEBS for collision avoidance. Thus, it has been an important question to be addressed: Do drivers place too much trust in and overreliance on AEBS for collision avoidance? The answer can be given by investigating the possibility of the overtrust and the overreliance with the theoretical framework proposed in this paper.

Since the system is activated only in cases of emergency, it would be very rare for an ordinary driver to see how the system works (i.e., chance-of-observation axis). It is thus highly possible that the driver will not be able to construct a precise mental model of the AEBS through the use of it. This suggests that it may be hard for the driver to engender a sense of trust in the system, especially in terms of "performance" (i.e., dimension-of-trust axis). What happens then? No possibility

for the driver to place overtrust in the AEBS? The answer may be negative. It is known that people may place inappropriate trust (i.e., overtrust), especially in terms of process and/or purpose dimensions, without having any concrete experience or evidence proving that the object is trustworthy; see, for example, [17]. Our experience with drivers' overtrust in and overreliance on airbag systems also supports the concern with overtrust in AEBS.

Suppose the driver places overtrust in the system. Does that mean that the driver relies on the system too much (i.e., overreliance)? In one sense, the answer may be positive. Itoh et al. [26] conducted a driving simulator experiment and found that drivers shortened the time headway while they followed a lead vehicle when an AEBS for collision avoidance was available. This is an example of commission-like actions. However, such overreliance was partly due to repetitive experience of the AEBS in the experiment. In reality, it is rare for a driver to observe cases where the automatic collision avoidance brake is activated.

In addition, the drivers were not distracted at all in the experiment even when the AEBS was available, and the drivers' reaction against the rapid deceleration of the lead vehicle was not delayed in the experiment of Itoh et al. [26]. That is, the drivers may not rely on the AEBS excessively in a sense that they allocate their resources to something else at the risk of their life (i.e., benefit-expected axis).

In case of an ADAS designed for use in normal driving situations, even if the system's behavior was not what the driver expected, there would be enough time for the driver to override the system to cope with the circumstances himself or herself. However, in case of an ADAS for emergency use, even if the driver noticed that the system's behavior was not what he or she expected, no time may be left for him or her to correct it (i.e., time-allowance-for-human-intervention axis).

The above discussion suggests that AEBS for collision avoidance could be free from drivers' overreliance if the system is designed appropriately.

In Japan, a national advanced safety vehicle (ASV) project discussed this issue. One of the authors of this paper was the leader of the task force in the ASV project to investigate design requirements for AEBS with collision avoidance functionality. As a conclusion, the ASV task force approved that the AEBS may be developed as a collision "avoidance" system, instead of a collision damage mitigation system. Such collision avoidance AEBS may not interfere with the driver's own actions (by letting it apply the automatic brakes at the latest time possible), but still it can avoid a collision against a forward obstacle effectively. Human factors viewpoints played major roles in determining the design requirements on the AEBS timing to initiate an automatic emergency braking and its deceleration rate. In fact, they were determined through the analyses of drivers' braking behaviors in normal and critical traffic conditions. Moreover, a couple of conventional requirements for the AEBS were abolished from human factors viewpoints (e.g., to reduce mode confusion or automation surprises). Based on the conclusion of the ASV task force, the Ministry of Land, Infrastructure, and Transport has been revising the design guidelines for the AEBS.

5. Discussions

This paper has proposed a theoretical framework to discuss the driver's overtrust in and overreliance on ADAS in a precise manner. Overtrust and overreliance are distinguished rigorously, and their characteristics are illustrated by introducing some viewpoints (or aspects and axes). It has been shown that our theoretical frame enables precise description, classification, rigorous analysis, and evaluation of the driver's overtrust in and overreliance on ADAS. Since the framework distinguishes the target object of the driver's overtrust, it can be used to derive a countermeasure for reducing the possibilities of the driver's overtrust. In other words, a systematic investigation can be made possible to determine whether overtrust in question may be alleviated by improving human-machine interface, or by preparing a better operation manual, or by providing the drivers with opportunities to acquire knowledge and/or to improve skills, or by some other means.

It would be apparent that alleviation or prevention of overtrust in or overreliance on the ADAS and its effects on degradation of safety of the car-driver system are closely linked to the issue of authority and responsibility. It is sometimes useful to provide the driver with multilayered assist functions [1]. In the first layer, a driver's situation recognition and understanding are enhanced for proper situation diagnostic decisions and associated action selection decisions. In the second layer, the ADAS monitors the driver's behaviours and traffic conditions to evaluate whether his or her intent and behaviours match the traffic conditions. When the ADAS detects a deviation from normality (for instance, by detecting behaviours or postures that suggest the driver's overtrust or its resulting overreliance), it gives the driver an alert to make him or her return to normality. In the third layer, the ADAS provides the driver with automatic safety control functions, if the deviation from normality still continues to be observed or if little time is left for the driver to cope with the traffic conditions. The situation-adaptive ADAS adjusts its assist functions dynamically, so that they may fit to the human's intent, psychological/physiological conditions, and the traffic conditions. The adjustment of assist functions is made in a *machine-initiated* manner [27–29] by inferring intent and conditions of the human through monitoring his or her behaviours. It is proven mathematically in [29] that a machine-initiated trading of authority based on the machine's interpretation of the situation and the human's behaviour is indispensable for assuring safety of the car-driver system, although the machine-initiated policy is not human-centered in the sense of [4].

The driver's control action may be classified into three categories: (1) an action that needs to be done in a given situation, (2) an action that is allowable in the situation and thus it may either be done or undone, and (3) an action that is inappropriate and thus must not be done in the situation. Assuming sensing technology for the computer (ADAS), two states may be distinguished for each control action: (a) "detected," in which the computer judges that the driver is performing the control action, and (b) "undetected," in which the control action is not detected by the computer (Figure 2).

FIGURE 2: Control action in a given situation.

Case A represents a circumstance with the driver's omission-like action selection, while case B depicts a circumstance with the driver's commission-like action selection and implementation. These mismatches between the driver's action selection decision and the given situation can occur when the driver may place overreliance on the ADAS, as has been discussed already. Then the question becomes, "what is a sensible and effective countermeasure for the ADAS in such circumstances? Is it enough for the ADAS to set off an alert to let the driver resolve the mismatch himself or herself? Or, is it better for the ADAS to initiate an automatic control action to cope with the situation?" Inagaki and his colleagues have shown that the authority may be given to the ADAS, so that (i) it can take an automatic safety control action that the driver failed to perform, or (ii) it can take a protective action (*soft protection* or *hard protection*) that tries to prevent the driver's inappropriate action causing an accident or an incident [30–32].

References

[1] T. Inagaki, "Smart collaboration between humans and machines based on mutual understanding," *Annual Reviews in Control*, vol. 32, no. 2, pp. 253–261, 2008.

[2] E. Hollnagel and A. Bye, "Principles for modelling function allocation," *International Journal of Human Computer Studies*, vol. 52, no. 2, pp. 253–265, 2000.

[3] G. Klein, "A recognition-primed decision (RPD) model of rapid decision making," in *Decision Making in Action*, G. Klein, J. Orasanu, and R. Calderwood, Eds., pp. 138–147, Ablex, 1993.

[4] C. E. Billings, *Aviation Automation—The Search for a Human-Centered Approach*, LEA, 1997.

[5] T. Inagaki, "Design of human-machine interactions in light of domain-dependence of human-centered automation," *Cognition, Technology and Work*, vol. 8, no. 3, pp. 161–167, 2006.

[6] M. R. Endsley and E. O. Kiris, "The out-of-the-loop performance problem and level of control in automation," *Human Factors*, vol. 37, no. 2, pp. 381–394, 1995.

[7] N. B. Sarter and D. D. Woods, "How in the world did we ever get into that mode? Mode error and awareness in supervisory control," *Human Factors*, vol. 37, no. 1, pp. 5–19, 1995.

[8] R. Parasuraman and V. Riley, "Humans and automation: use, misuse, disuse, abuse," *Human Factors*, vol. 39, no. 2, pp. 230–253, 1997.

[9] N. B. Sarter, D. D. Woods, and C. E. Billings, "Automation surprises," in *Handbook of Human Factors and Ergonomics*, G. Salvendy, Ed., pp. 1926–1943, John Wiley & Sons, 2nd edition, 1997.

[10] T. Inagaki and J. Stahre, "Human supervision and control in engineering and music: similarities, dissimilarities and their implications," *Proceedings of the IEEE*, vol. 92, no. 4, pp. 589–600, 2004.

[11] R. Parasuraman, R. Molloy, and I. L. Singh, "Performance consequences of automation-induced 'complacency'," *International Journal of Aviation Psychology*, vol. 3, no. 1, pp. 1–23, 1993.

[12] K. L. Mosier, L. J. Skitka, S. Heers, and M. Burdick, "Automation bias: decision making and performance in high-tech cockpits," *International Journal of Aviation Psychology*, vol. 8, no. 1, pp. 47–63, 1998.

[13] J. Meyer, "Effects of warning validity and proximity on responses to warnings," *Human Factors*, vol. 43, no. 4, pp. 563–572, 2001.

[14] T. B. Sheridan and R. Parasuraman, "Human-automation interaction," in *Reviews of Human Factors and Ergonomics*, R. S. Nickerson, Ed., vol. 1, pp. 89–129, HFES, 2005.

[15] *Complacency*, Merriam-Webster Online Dictionary, 2010.

[16] N. Moray and T. Inagaki, "Attention and complacency," *Theoretical Issues in Ergonomics Science*, vol. 1, no. 4, pp. 354–365, 2001.

[17] J. Lee and N. Moray, "Trust, control strategies and allocation of function in human-machine systems," *Ergonomics*, vol. 35, no. 10, pp. 1243–1270, 1992.

[18] M. Itoh, "Toward overtrust-free advanced driver assistance systems," *Cognition, Technology, and Work*, vol. 14, no. 1, pp. 51–60, 2012.

[19] T. Inagaki, "New challenges on vehicle automation: human trust in and reliance on adaptive cruise control systems," in *Proceedings of the IEA (CD-ROM)*, p. 4, 2003.

[20] G. J. S. Wilde, *Target Risk*, PDE Publications, Toronto, Canada, 1994.

[21] OECD Road Transport Research, *Behavioral Adaptations to Changes in the Road Transportation System*, OECD, Paris, France, 1990.

[22] M. Weinberger, H. Winner, and H. Bubb, "Adaptive cruise control field operational test—the learning phase," *JSAE Review*, vol. 22, pp. 487–494, 2001.

[23] B. D. Seppelt and J. D. Lee, "Making adaptive cruise control (ACC) limits visible," *International Journal of Human Computer Studies*, vol. 65, no. 3, pp. 192–205, 2007.

[24] D. A. Dickie and L. N. Boyle, "Drivers' understanding of adaptive cruise control limitations," in *Proceedings of the 53rd Human Factors and Ergonomics Society Annual Meeting (HFES '09)*, pp. 1806–1810, October 2009.

[25] K. Tanno, M. Kohno, K. Ono et al., "Fatal cardiovascular injuries to the unbelted occupant associated with airbag deployment: two case-reports," *Legal Medicine*, vol. 2, no. 4, pp. 227–231, 2000.

[26] M. Itoh, Y. Fujiwara, and T. Inagaki, "Driver behavioural changes through interactions with an automatic brake systems," *Transactions of the Society of Instrument and Control Engineers*, vol. 47, no. 11, pp. 512–519, 2011 (Japanese).

[27] T. Inagaki, "Adaptive automation: sharing and trading of control," in *Handbook of Cognitive Task Design*, E. Hollnagel, Ed., pp. 147–169, LEA, 2003.

[28] M. W. Scerbo, "Theoretical perspectives on adaptive automation," in *Automation and Human Performance*, R. Parasuraman and M. Mouloua, Eds., pp. 37–63, LEA, 1996.

[29] T. Inagaki and T. B. Sheridan, "Authority and responsibility in human-machine systems: probability theoretic validation of machine-initiated trading of authority," *Cognition, Technology & Work*, vol. 14, no. 1, pp. 29–37, 2012.

[30] T. Inagaki, M. Itoh, and Y. Nagai, "Efficacy and acceptance of driver support under possible mismatches between driver's intent and traffic conditions," in *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society (HFES '06)*, pp. 280–283, October 2006.

[31] T. Inagaki, M. Itoh, and Y. Nagai, "Driver support functions under resource-limited situations," in *Proceedings of the 51st Annual Meeting of the Human Factors and Ergonomics Society (HFES '07)*, pp. 176–180, October 2007.

[32] T. Inagaki, M. Itoh, and Y. Nagai, "Support by warning or by action: which is appropriate under mismatches between driver intent and traffic conditions?" *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E90-A, no. 11, pp. 2540–2545, 2007.