In P.C. Cacciabue et al (eds.), Human Modelling in Assisted Transportation. Springer, Milan, pp 33-41, 2011

To What Extent May Assistance Systems Correct and Prevent 'Erroneous' Behaviour of the Driver?

Toshiyuki INAGAKI

University of Tsukuba

Department of Risk Engineering

Tsukuba 305-8573 Japan

inagaki@risk.tsukuba.ac.jp

Abstract

An error in situational recognition may occur while driving a car, and the error can sometimes result in an 'erroneous' behaviour of the driver. Whether the driver assistance system can cope with such a circumstance depends on to what extent the authority is given to the system. This paper discusses the need of machine-initiated authority trading from the driver to the assistance system for assuring driver safety. A theoretical framework is also given to describe and analyze the driver's overtrust in and overreliance on such a driver assistance system.

1 Introduction

Main topics of the HMAT Workshop include "methods and tools to prevent erroneous behaviour to mitigate its consequences." Driving a car requires a continuous process of perception, cognition, action selection, and action implementation. An error in situational recognition may occur while driving a car, and the error can sometimes result in an 'erroneous' behaviour of the driver. In order to "prevent erroneous behaviours" of car drivers, it is most fundamental to provide the drivers with assistances for perception and cognition so that the drivers can grasp a situation clearly and correctly. Once the situation is properly understood, it is usually straightforward for the humans to determine what actions need to be done in the situation [5, 13]. Design of human-machine interfaces based on onboard self-sensing technology as well as vehicle-to-vehicle and vehicle-to-infrastructure communication technologies play important roles in implementing assistance

functions to enhance, augment, and complement driver capabilities for perception and cognition.

What if an error in situational understanding has occurred in spite of such assistances for perception and cognition and if an ‘erroneous’ behaviour of the driver has been detected? A natural action for the driver assistance system would be to set off warnings to urge the driver to stop or correct the ‘erroneous’ behaviour. Warnings are expected to assist the driver’s action selection.

Suppose the driver did not respond to the warnings. Does the assistance system perform nothing but observe consequence of the driver’s ‘erroneous’ behaviour to occur? Or, may the assistance system take some control action to avoid such a consequence? Answers to the questions are not so simple. When the control action is not directed by the driver but is decided by the assistance system, an issue of authority and responsibility arises, because the driver is assumed to be always in charge and command: The Convention of Road Traffic [3], for instance, states that “Every driver of a vehicle shall in all circumstances have his vehicle under control so as to be able to exercise due and proper care and to be at all times in a position to perform all manoeuvres required of him” (Article 13.1).

This paper investigates the issue of authority and responsibility between the driver and the assistance system, and argues that the assistance system may be allowed to trade authority from the driver to the assistance system based on its decision for assuring safety. When the assistance system is capable to correct and prevent ‘erronoues’ behaviour of the driver, overtrust in and overreliance on the assistance system become an important issue: Regulatory authorities often express their concerns over the possibility of the drivers’ behavioural changes in which they place excessive trust in and reliance on the driver assistance systems [8]. This paper gives a theoretical framework for discussing the driver’s overtrust in and overreliance on autonomous assistance systems in a rigorous manner.

2 Authority and Responsibility

‘Erroneous’ behaviours may be classified into two types: (1) omission-like behaviour failing to select or implement an action needed in a given situation and (2) commission-like behaviour to select and implement an action inappropriate to a given situation. The former corresponds to case A and the latter to case B in Fig.1, respectively, under the assumption of technology to sense and interpret traffic conditions and driver behaviours, as well as the three-class categorization of the driver’s control action as (a) an action that needs to be done in the given situation, (b) an action that is allowable in the situation, and (c) an action that is inappropriate and thus must not be done in the situation.

Suppose the driver assistance system has determined that the driver’s behaviour is ‘erroneous.’ The assistance system must determine which is more sensible and effective in the circumstance, a warning type support in which a warning is set off

to urge the driver to react to the situation, or an action type support in which the assistance system executes an autonomous safety control action?

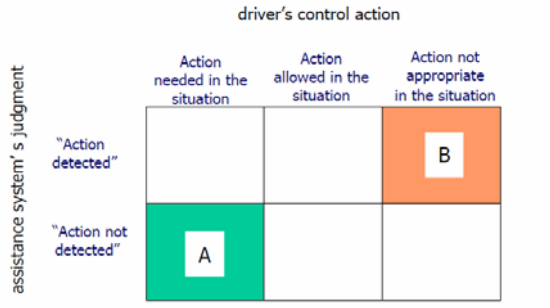

Fig. 1 Control action in a given situation

Consider first characteristics of the warning type support. If the 'erroneous' behaviour is of an omission-like type (case A), the warning directs the driver to implement at once a necessary but missing action. If the 'erroneous' behaviour is of a commission-like type (case B), the warning tries to tell the driver to stop doing the inappropriate action. In either case, the driver is maintained as the final authority over the assistance system; it is the driver who decides whether to accept and implement what was meant by the warning. The relation between the driver and the assistance system is fully compatible with the Convention on Road Traffic and the *human-centered automation* principles claiming that the human bears the ultimate responsibility for safety and therefore the human must be in command; see, e.g., [1, 2, 6]. In fact the assistance system's situation understanding can be incorrect because of its limitation. At the same time, the human-centeredness of the warning type support can fail to assure the driver safety: The driver may not be able to cope with the situation, because of a short time allowance or because of internal/external distractions. It can also happen that the driver disregards a given warning based on a 'reasonable' but wrong interpretation of the warning [12].

Consider next characteristics of the action type support. If the 'erroneous' behaviour is of an omission-like type (i.e., case A in Fig. 1), the assistance system executes an action that the driver failed to perform. If the 'erroneous' behaviour is of a commission-like type (i.e., case B in Fig. 1), the assistance system applies control to prohibit the driver to continue doing the inappropriate action. In either case, the authority is traded from the driver to the assistance system, and it is the assistance system that determines and implements the authority trading, which is sometimes called *machine-initiated automation invocation* [10]. Thus the relation between the driver and the assistance system is not fully compatible with either the Convention on Road Traffic or the *human-centered automation* principles. However, as long as the human has limitation, there is a space for the assistance system to execute a control action on behalf of the driver or to correct the driver's 'erroneous' action.

In the design of a mechanism for machine-initiated automation invocation, it is useful to distinguish *hard protection* and *soft protection*. In hard protection, the driver is not allowed to override the assistance system's control action. In soft protection, on the other hand, the driver is given authority to override the control action applied by the assistance system. It is sometimes observed that the drivers prefer soft protection to hard protection, although the soft protection may not be perfect in preventing the driver's 'erroneous' action [11, 12]. The assistance system with a mechanism for machine-initiated automation invocation gives the driver a slight chance to behave as the final authority over the automation, when the design of the assistance system is of soft protection type.

3 Advanced Safety Vehicle: A Japan's National Project

Advanced Safety Vehicle (ASV) is a car equipped with technology-based driver assistance systems to enhance safety under normal as well as time-critical situations. The ASV project has been conducted since 1991 under the cooperation of industries, academia, and the government. It is assumed there that the driver must be always in charge and that the driver assistance systems are allowed to provide the driver with 'assistance'. Some guidelines for designing driver assist systems are: (1) The system should act in line with intent of the driver. (2) The system should assist the driver to perform safe driving and steady operation. (3) The driver should monitor operations of the assist system when it is in action. (4) The system should not cause overconfidence or overtrust of the driver. (5) The system, when it is in action, should allow the driver's intervention to override its operation. (6) The system's control should be smoothly passed over to the driver when the situation goes beyond the range of the system [7, 16]. The design principles and guidelines for the driver assistance systems were discussed and established in the second 5-year phase of the project (1996-2000) through investigations of negative effects of automation, such as the out-of-the loop performance problem, loss of situational awareness, overtrust, distrust, and automation surprises; see, e.g., [4, 9, 18, 20, 21, 23].

The ASV project has developed various systems that provide the drivers with assistances for perception, cognition, and action selection. However, the Ministry of Land, Infrastructure and Transport (MLIT) as well as National Police Agency of the Government of Japan have been taking a cautious stance on putting systems into practical use when the assistance systems are for action implementation. It is true, of course, that there are such systems. The adaptive cruise control (ACC) and the lane keeping assistance (LKA) are examples of systems for assisting driver's action implementation by *relieving* the driver's load. The electronic stability control (ESC) and the antilock brake system (ABS) are also examples of systems for assisting driver's action implementation by *amplifying* or *extending* the capabilities of the driver.

The arguments become different when it comes to the assistance systems that have capabilities to *back up* or *replace* the driver. Take, as an example, the pre-crash safety (PCS) system that is sometimes called the advanced emergency braking system (AEBS). When the host vehicle is approaching relatively fast to a lead vehicle, the PCS firstly tightens the seat belt and adds a warning to urge the driver to put on the brake. If the PCS determined that the driver is late in braking, then it applies the brake automatically based on its decision. However, the PCS is currently implemented as a *collision damage mitigation system*, in stead of as a *collision avoidance system*. Behind the design decision to 'downgrade' the PCS, there has been concern among the regulatory authorities that "If a driver assistance system would perform every safety control action automatically, the driver may become overly reliant on the assistance system, without paying attention to the traffic situations himself or herself."

Although the above 'concern' seems to be reasonable, there have been some discussions in the ASV project that more precise investigations would be necessary so as not to lose opportunities for the drivers (especially, elder drivers) to be benefited by the assistance system that may back up or replace them when appropriate. The next two sections give a theoretical framework to describe and analyze overtrust in and overreliance on the driver assistance system. Although the two terms 'overtrust' and 'overreliance' are often used as if they are synonyms, this paper differentiates them rigorously.

4 Overtrust

Overtrust in a driver assistance system is an incorrect diagnostic decision to conclude that the assistance system is trustworthy, when it actually is not. This section gives two axes for discussing overtrust in the assistance system. The first axis is the *dimension of trust* and the second the *chance of observations*.

The first axis is to describe in which way the driver can overrate trust. Lee and Moray [14] distinguished four dimensions of 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. Three types of overtrust can be distinguished depending on which dimension among (b) through (d) is overrated; the first dimension (a) is usually met in cases of the driver assistance systems.

Overrating of (b) can be seen in a case where a driver thought, "The assistance system has been responding perfectly to all the events that I have encountered so far. Whatever events may occur, the system will take care of them nicely." Improper evaluation of (c) is seen in a case where a driver has been using an assistance system without reading the user's manual at all by thinking, "It would be

quite alright even if I do not know the details of the system functions." Overestimation of (d) may be seen in a case where a driver believes that "I do not understand why my assistance system is doing such a thing. However, it must be doing what it thinks it necessary and appropriate."

The second axis for investigating overtrust is to describe how often the driver can see the assistance system functions. The chance of observations affects the ease of constructing a mental model of the assistance system. The possibility of the driver's overtrust can differ depending on whether the assistance system is for use in normal driving or is for use in emergency.

Take the ACC as an example of the assistance system to reduce the driver workload in normal driving. Based on a large number of opportunities to observe the ACC's functioning repeatedly in daily use, it would be easy for the driver to construct a mental model of the ACC. If the driver has been satisfied with 'intelligent' behaviours of the ACC, it may be natural for him or her to place trust in the assistance system. However, the trust can sometimes be overtrust. Suppose the driver encounters a new traffic condition that is seemingly similar to a previous one but is slightly different. If the driver expected that the ACC would be able to cope with the situation without any intervention of the driver, it can be an overestimation of the ACC's functionality.

Take next the PCS as an example of the assistance system activated only in emergency to assure the driver safety. It would be rare for an ordinary driver to see the PCS works, and he or she may not be able to construct a complete mental model of the PCS because of lack of enough number of chances to experience the PCS. The driver might have been told (by a car dealer, for instance) that the PCS shall be activated automatically in emergency. However, the driver may not be fully convinced because of lack of chances to observe himself or herself that the PCS works properly and constantly when necessary.

5 Overreliance

Overreliance on a driver assistance system is an incorrect action selection decision determining to rely on the assistance system by placing overtrust in it. Regarding overreliance on automated warning systems, there are relevant studies in aviation domain; see, e.g., [15, 17, 19, 22]. Suppose that the automated warning system almost always alerts the human when an undesirable event occurs. Although it is possible for a given alert to be false, the human can be confident that there is no undesirable event as long as no alert is given (A similar situation can happen in automobile domain when the driver is provided with a communication-based alert from the road infrastructure to let the driver know of an approach or existence of cars on a crossing road behind some buildings). Meyer [15] used the term 'reliance' to express such a response of the human. If the human assumed that the automated warning system will always give alerts when an undesirable event oc-

curs, that may be overtrust in the warning system and the resulting reliance on the warning system is overreliance. The definition of overreliance on the driver assistance system, given at the beginning of this section, is a generalization of that of overreliance on the warning system in the previous studies in the sense that the assistance system is not only for setting off warnings but also for executing control actions.

Two axes are given for overreliance in the assistance systems. The first axis is the *benefits expected* and the second the *time allowance for human intervention*.

The first axis is to describe whether the driver can produce some benefits by relying on the assistance system. 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 finding 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.

The discussion can be quite different in case of PCS. The PCS is activated only in emergency, and the time duration for the PCS 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 PCS, to something else to produce extra benefit within the several seconds. A similar argument may apply to other assistance systems designed for emergency.

The second axis, time allowance for human intervention, 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. In case of ACC, it may not be hard for the driver to intervene to override the ACC when its performance was not satisfactory. However, in case of PCS, it might be unrealistic to assume that the driver can intervene into control by the PCS when he or she decided that the PCS's performance was not satisfactory, because the whole process of monitoring and evaluation of PCS's performance as well as decision and implementation of intervention must be done within a few seconds.

6 Concluding Remarks

It is often useful to provide the driver with multi-layered assistance functions [7]. In the first layer, driver's perception and situation recognition are enhanced to lead to proper situation diagnostic decisions and associated action selection decisions. In the second layer, the assistance system monitors the driver's behaviours as well as traffic conditions to evaluate whether his or her intent and behaviours match the traffic conditions. When the assistance system has detected a *deviation from normality*, it gives the driver an alert to make him or her return to normality. In the third layer, the assistance system 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 situation. In such a *situation-adaptive assistance system*, a mechanism is needed to decide and implement authority trading in a machine-initiated manner, which poses an issue of authority and responsibility [7, 10].

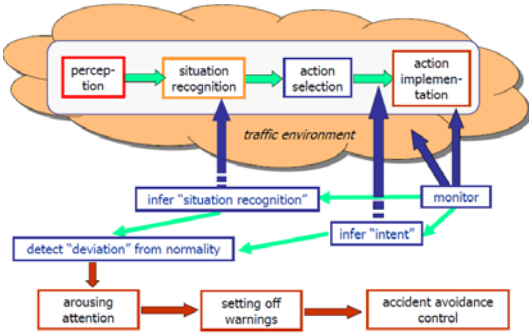

Fig 2 Driver monitoring and situation-adaptive assistance

The issue is further linked to that of the driver's overtrust in and overreliance on the assistance system. Actually, there is a serious concern that the driver may place overreliance on an autonomous and smart driver assistance system. This paper has given a general framework for describing overtrust in and overreliance on the assistance system, and has argued that whether the driver puts overtrust in or overreliance on the assistance system can vary depending on the characteristics of the assistance system. Based on the framework, the following argument may be possible for PCS, as an example: "Since the PCS 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 hard for the driver to construct a precise mental model of the PCS, and may be hard for him or her to engender a sense of trust in the system (i.e., dimension-of-trust axis). However, it is known that people may place inappropriate trust (i.e., overtrust) without having any concrete evidence proving that the object is trustworthy. Now, let us assume that the driver places overtrust in the assistance system. We have to ask whether the driver may rely on the system excessively (i.e., overreliance). In case of PCS, even if the driver noticed that the system's behavior was not what was expected, no time may be left for the driver to intervene and correct it. In spite of that, does the driver rely on the PCS (i.e., overreliance) and allocate his or her resources to something else at the risk of his or her life? The answer would be negative."

A task force was set up in December 2009 in the ASV project to investigate sharing and trading of authority and responsibility between the driver and the assistance systems, as well as the driver's overtrust in and overreliance on the assistance systems. Multi-disciplinary analyses and discussions, including legal aspects, are planned in the task force. It is expected to draw guidelines for designing driver assistance systems of next generation.

References

[1] Billings CE (1997) Aviation automation – the search for a human-centered approach. LEA, Mahwah
[2] Cacciabue PC (2004) Guide to applying human factors methods: human error and accident management in safety critical systems. Springer, Berlin
[3] Convention on Road Traffic (1968) 1993 version & amendments in 2006
[4] Endsley MR, Kiris EO (1995) The out-of-the-loop performance problem and the level of control in automation. Hum Factors 37(2):3181-3194
[5] Hollnagel E, Bye A (2000) Principles for modeling function allocation. Int J Human-Comp Stud 52:253-265
[6] Inagaki T (2006) Design of human-machine interactions in light of domain-dependence of human-centered automation. Cogn Tech Work 8(3):161-167
[7] Inagaki T (2008) Smart collaborations between humans and machines based on mutual understanding. Annual Reviews in Control 32:253-261
[8] Inagaki T (2010) Traffic systems as joint cognitive systems: issues to be solved for realizing human-technology coagency. Cogn Tech Work. doi:10.1007/s10111-010-0143-6
[9] Inagaki T, Stahre J (2004) Human supervision and control in engineering and music: Similarities, dissimilarities, and their implications. Proc IEEE 92(4):589-600
[10] Inagaki T, Sheridan TB (2008) Authority and responsibility in human-machine systems: Is machine-initiated trading of authority permissible in the human-centered automation framework? In: Proceedings of applied human factors and ergonomics 2008 (CD-ROM) 10 p
[11] Inagaki T, Itoh M, Nagai Y (2007) Support by warning or by action: Which is appropriate under mismatches between driver intent and traffic conditions? IEICE Trans Fundam E90-A(11):264-272
[12] Inagaki T, Itoh M, Nagai Y. (2008) Driver support functions under resource-limited situations. J Mechanical Systems for Transportation and Logistics, 1(2): 213-222
[13] Klein G. (1993) A recognition-primed decision (RPD) model of rapid decision making. In: Klein G et al (eds) Decision making in action. Ablex, New York, pp 138-147
[14] Lee, J.D. & Moray, N. (1992) Trust, control strategies and allocation of function in human machine systems. Ergonomics 35(10):1243-1270
[15] Meyer J (2001) Effects of warning validity and proximity on responses to warnings. Hum Factors 43(4):563-572
[16] MLIT (2007) ASV; the bridge to an accident-free society. Ministry of Land, Infrastructure and Transport, Government of Japan
[17] Mosier K, Skitka LJ, Heers S, Burdick M (1998) Automation bias: decision making and performance in high-tech cockpits. Int J Aviation Psychol 8:47-63
[18] Parasuraman R, Riley V (1997) Humans and automation: use, misuse, disuse, abuse. Hum Factors 39(2):230-253
[19] Parasuraman R, Molloy R. Singh IL (1993) Performance consequences of automation-induced 'complacency.' Int J of Aviation Psychol 3(1):1-23
[20] Sarter NB, Woods DD (1995) How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Hum Factors 37(1):5-19
[21] Sarter NB, Woods DD, Billings CE (1997) Automation surprises. In: Salvendy G (ed) Handbook of human factors and ergonomics, 2nd edn. Wiley, New York, pp 1926-1943
[22] Sheridan TB, Parasuraman R (2005) Human-automation interaction. In: Nickerson RS (ed) Reviews of human factors and ergonomics, Vol 1, Human Factors and Ergonomics Society, Santa Monica, pp 89-129
[23] Wickens CD (1994) Designing for situation awareness and trust in automation. In: Proceedings of IFAC integrated systems engineering, pp 77-82.