15th Triennial World Congress, Barcelona, Spain

# DYNAMICS OF DRIVERS' TRUST IN WARNING SYSTEMS

**G. Abe*[1], M. Itoh** and K. Tanaka****

* *Safety and IT Research Division, Japan Automobile Research Institute*
*Tsukuba 305-0822, Japan*
** *Graduate School of Information Systems, University of Electro-Communications*
*Chofu 182-8585, Japan*

Abstract: This paper focuses on how the behavior of drivers in response to warning systems is influenced by their trust in those warning systems. The experimental examinations show that a) false alarms result in a reduction in trust and an increased reaction time to braking, b) a lack of alarms result in a delay in the recognition of hazardous situations, c) excessive trust induces significant delays in such situations and d) a lack of alarms more significantly reduces trust than false alarms do. *Copyright © 2002 IFAC*

Keyword: Alarm system, Driver behavior, Reliability, Safety, Transportation

## 1. INTRODUCTION

Decreasing road traffic collisions is accepted as an very important objective throughout the world. As humans make mistakes when driving, it is necessary to assist drivers to achieve safe efficient driving. In recent years a number of new technologies, known as intelligent transportation systems (ITS) , have been developed in order to reduce incidents and to improve safety. These technologies include collision-warning and collision-avoidance systems which contribute to mitigate a delay in the recognition of hazardous situation caused by drivers' inattention. However, these devices may fail or behave unpredictably and for such systems to operate effectively, it is important for drivers and machines to perform their separate tasks appropriately. The degree of cooperation between humans and machines can have a dramatic influence on the performance of the ITS.

Previous research has shown that trust in machines is a key factor in attaining appropriate cooperation between humans and machines (Muir, 1994; Muir and Moray, 1996; Riley,1996) . Moreover, trust often determines machines usage (Parasuraman and Riley, 1997; Lee and Kantowitz, 1998) . Humans may not use a reliable machine that is believed to be untrustworthy. Conversely, a machine that is believed to be trustworthy may be relied upon even when it malfunctions. Trust changes dynamically (Lee and Moray, 1992; Muir and Moray, 1996) and has three dimensions, namely predictability, dependability and faith that alter according to humans' trial-and-error-experiences (Muir, 1987) . Moray, et al. (2000) have shown that a main factor in the development of trust is the reliability of machines. Itoh, et al. (1999) have shown that the dynamics of trust depends not only on the reliability of machines but also

[1] Now the author is a doctoral canditate in Research School in Ergonomics and Human Factors, Loughborough University,UK.

on the occurrence patterns of machine malfunctions.

Thus drivers' behavior toward warning systems may alter according to their trust in those systems. This paper examines how, as a result of influencing drivers' trust in systems, malfunctions reduce the efficiency of warning systems from the viewpoint of a relationship between drivers' trust and behavior. Moreover two types of malfunctions namely false alarms and a lack of alarms are considered in this study. Because each malfunction presents different phenomena to drivers, each malfunction influences drivers' trust in warning systems differently. False alarms which produce unnecessarily warnings and frequent false alarms result in a reduction in drivers' trust in warning systems. Accordingly, drivers may ignore the alarms or temper their reactions to them. Conversely a lack of alarms means that alarms do not issue even though the warning criterion for each systems is fulfilled. Therefore a lack of alarms may result in the absence of any appropriate risk avoidance action when drivers trust warning systems excessively.

This article addresses three issues concerning the efficiency of warning systems used in a driving simulator. First, the way in which a reduction of trust caused by false alarms affects driver reaction time to alarms is investigated. Second, how response time to risk avoidance action is influenced by a lack of alarms is investigated. Finally, the dynamics of trust under conditions of both false alarms and a lack of alarms in the same warning system are investigated.

## 2. DRIVING SIMULATOR MODEL

### *2.1 Drivers' Task*

For this research The Driving Simulator owned by JARI (Japan Automobile Research Institute) was used. The road conditions comprised a straight freeway, and an absence of pedestrians and vehicles except a leading vehicle and a following vehicle. The leading vehicle travels at 80km/h. The driver under investigation is given two tasks. One is to keep his/her own vehicle speed at 80km/h. The other is to avoid risk. The driver must take appropriate action when the leading vehicle decreases its speed.

### *2.2 Rear-End Collision Warning System*

This research uses a rear-end collision warning system, a new technology which has recently entered the marketplace. This system contributes to the avoidance of rear-end collisions caused by driver inattentions. In this research, a Stop-Distance-Algorithm (SDA) is used as one of the trigger logics for the warning (Wilson, et al., 1997). The SDA is based on "stop-distance", defined as the distance between the point at which braking begins and the point of rest of a leading vehicle and a following vehicle (Fig.1). A warning distance is then calculated based on the distance required for the leading vehicle to come to a complete halt; the warning would be triggered when the target vehicle's stopping distance exceeded this value (Fig.2). This equation includes three parameters, namely RT (reaction time, the assumed delay of drivers' reaction to an alarm), Df (assumed deceleration of the following vehicle), and Ds (assumed deceleration of the leading vehicle). The current conditions are defined as 1.0s, 5.88 $m/s^2$ and 5.88 $m/s^2$, respectively. The appropriate conditions were determined in preliminary experiments.

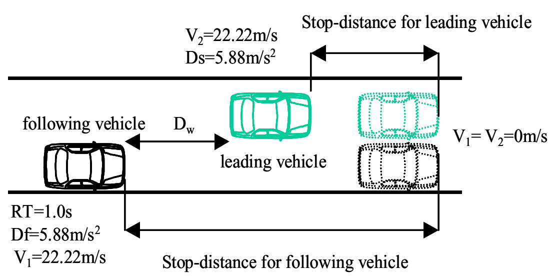


Fig. 1. Concepts of stop-distance for warning sysytem

$$D_w = V_1 \cdot RT + V_1^2 / 2 \cdot Df - V_2^2 / 2 \cdot Ds$$

$D_w$ : warning distance (m)

$V_1$ : velocity of following vehicle (m/s)

$V_2$ : velocity of leading vehicle (m/s)

$RT = 1.0s$

$Df = 5.88 m/s^2$

$Ds = 5.88 m/s^2$

Fig. 2. Defined equation of warning distance

### 2.3 *Malfunctions of Rear-End Collision Warning System*

In this study, two types of malfunctions, namely false alarms and a lack of alarms are assumed. A false alarm is defined as an alarm that issues in the absence of an imminent rear-end collision; a lack of alarms is defined as the failure of an alarm to issue when the leading vehicle decreases its speed sufficiently to fulfill the criteria for setting the warning trigger in this system.

## 3. EXPERIMENT I

### 3.1 *Experimental Conditions*

Eight JARI employees (5 males and 3 females) participated in this study. Each driver completed 15 trials. Each trial took about 3 minutes. During each trial, the leading vehicle decreased its speed rapidly once. The moment of speed reduction was always random. In this experiment, false alarms occurred in trials 7 to 9. In the other trials, the system worked correctly. The experiment consisted of guidance, a training session, data collection, and a post-test session (Fig. 3). All the same auditory alarms were presented to drivers during this experiment. The following data were collected as a measure of performance.

(1) The warning reaction time: this is the time period from the beginning of the issue of warning to the application of brakes by the driver (Fig. 4). This reaction time would be short if drivers believe the warning system to be trustworthy. On the other hand the reaction time would increase if drivers' trust in the warning system is vague or variable. Thus this measure may objectively estimate drivers' trustworthiness in the warning system.

(2) Subjective rating of trust in the rear-end collision warning system, measured using a standard 10-point rating scale used in studies on trust.

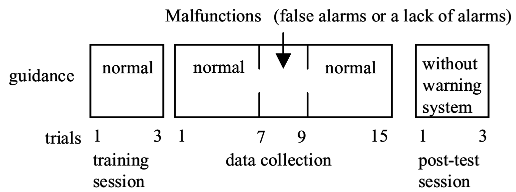


Fig. 3. Procedure for expriment

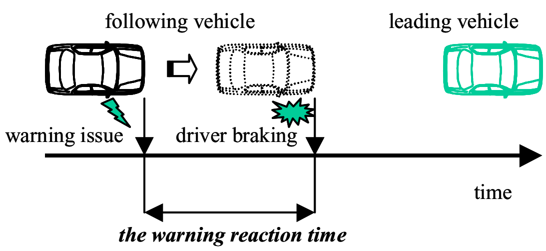


Fig. 4. The warning reaction time as a measure of proformance

### 3.2 *Results*

*The warning reaction time*. The solid line in Fig. 5 shows the average reaction time for each driver during each trial. In order to clarify the effect of false alarms on reaction time, all trials were divided into 3 blocks. Block1 includes the first 6 trials in which the warning system works correctly. Block2 includes trials 7 to 9 in which the warning system malfunctions. Block3 includes the last 6 trials in which the warning system works correctly. The dotted line in Fig. 5 shows the average reaction time for each block. These results show that the reaction times for Block2 are the longest; additionally, reaction times for Block3 do not return to the pre-malfunction levels of Block1. A one-way ANOVA on the warning reaction time among blocks reveals a significant main effect of blocks ($F(2,113)=6.11, p=0.003$). Tukey's HSD test shows significant differences in the warning reaction time between Block1 and Block2 ($p<0.01$), and between Block1 and Block3 ($p<0.01$).

*Trust in warning system*. The solid line in Fig. 6 shows the average subjective rating of trust for each trial. In order to clarify the effect of false alarms on drivers' trust, all trials were divided into 3 blocks as above. The dotted line in Fig. 6 shows the average subjective rat-

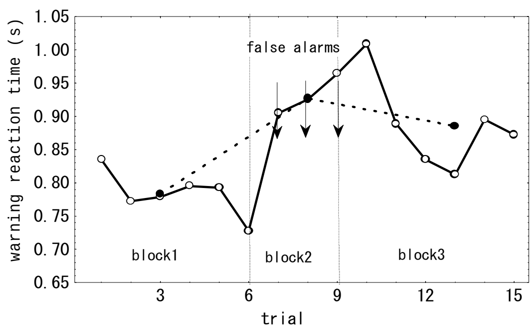


Fig. 5. Dynamics of warning reaction time

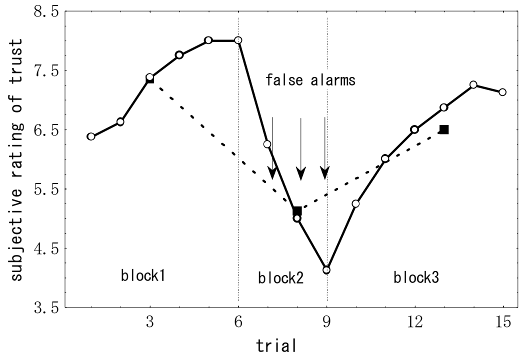

Fig. 6. Subjective rating of trust

ing of trust for each trial. Thus subjective ratings of trust in Block2 become significantly reduced, and that in Block3, the trust in the warning system does not recover to the level observed in Block1. A one-way ANOVA on subjective rating of trust among blocks revealed a significant effect of blocks ($F(2,117)=17.32$, $p<0.01$). Tukey's HSD test shows that there are significant differences in subjective rating of trust between Block1 and Block2 ($p<0.01$), and between Block1 and Block3 ($p<0.01$).

## 4. EXPERIMENT II

The problem of false alarms was discussed in experiment I, and their effects on both trust and warning reaction time is clear. Experiment II considers the effect of a lack of alarms on drivers' trust in warning systems, and driver behavior.

### *4.1 Experimental Conditions*

Nine JARI employees (5 males and 4 females) participated in this experiment. They had not participated in experiment I. The method of this experiment was the same as experiment I. The following data were collected as a measure of performance.

(1) The braking reaction time: this is the time period from the beginning of emergency deceleration of the leading vehicle to the application of brakes by the driver (Fig. 7). The braking reaction time indicates the drivers' awareness of emergency conditions.

(2) The delay of braking reaction time: this is defined as the difference between the average braking reaction time

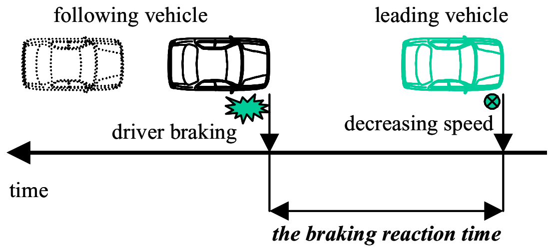

Fig. 7. The braking reaction time as measure of performence

for trials 1-5 (i.e. before any a lack of alarms are experience) and the average braking reaction time without the warning system. The braking reaction time without the warning system is measured three times in the post-test session (Fig. 3). For drivers who rely heavily on the warning system, this measure may become long.

(3) Subjective ratings of trust in rear-end collision warning systems: This measure is the same as experiment I.

### *4.2 Results*

*The braking reaction time.* The solid line in Fig. 8 shows the average braking reaction time for each trial. In order to clarify the effect of a lack of alarms on reaction time, all trials were divided into 3 blocks as in experiment I. The dotted line representing the average of braking reaction time in Fig. 8 shows that the reaction time for Block2 is the longest. A one-way ANOVA on the braking reaction time among blocks shows a significant effect of blocks ($F(2,132)=22.84, p<0.01$).

Fig. 9 illustrates the average braking reaction time for trials 7 to 9, and the average reaction time without a warning system (from the post-test session). In the absence of the warning system, drivers rely only upon their senses to avoid risk. Fig. 9 suggests that the brak-

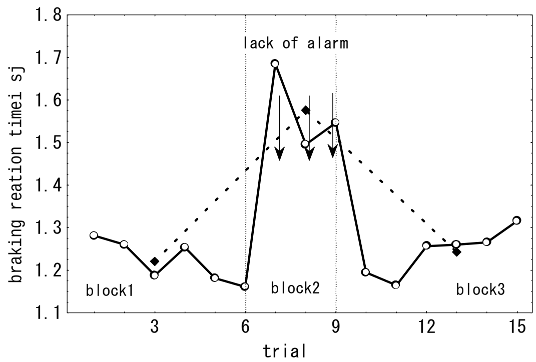

Fig. 8. Dynamics of braking reaction time

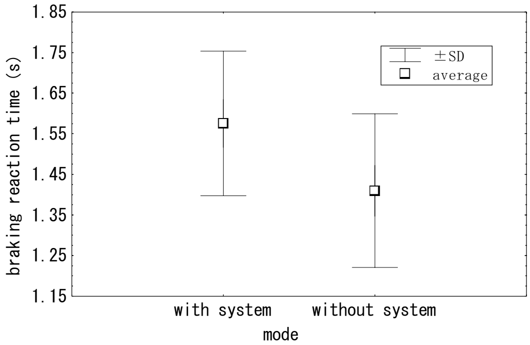


Fig. 9. The braking reaction time with warning system and without warning system

ing reaction time with the warning system installed but not operating is significantly higher than the braking reaction time with no warning system installed ($t(8)=5.10$, $p<0.01$).

*The length of delay and trust*. It was clear that a lack of alarms lengthened reaction time. Here, it is further argued that a relationship exists between the level of drivers' trust in warning system and the delay in braking reaction time. In Fig. 10, the X-axis and the Y-axis indicate the delay of braking reaction time and level of trust in the warning system, respectively. "Level of trust" is calculated as an average of the subjective ratings of trust for trials 1 to 6. Each plotted point represents one data for one driver. This figure shows an interesting relationship between level of trust and the delay of braking reaction time. For drivers who hold respectively low trust in the warning system, the delay in braking reaction time is short (solid circle in Fig. 10). Conversely for drivers who hold respectively high trust in the warning system, the delay in reaction time is greater (dotted circle in Fig. 10). In this experiment only one datum shows that for the driver who does not holds respectively high trust in warning systems, the delay in reaction time is

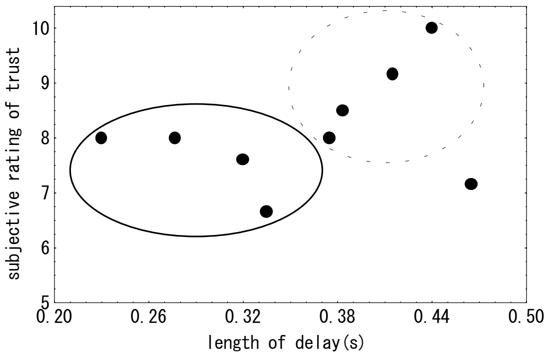


Fig. 10. Relationship between length of delay and level of trust

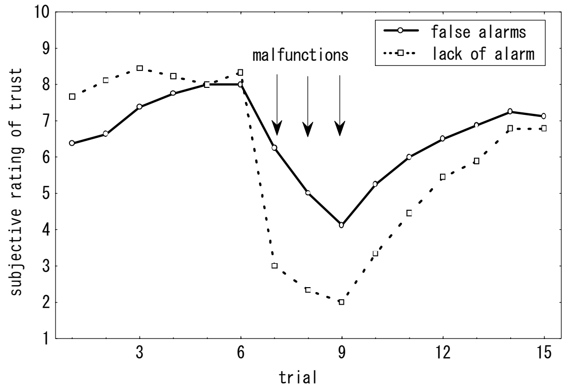


Fig. 11. Two types of dynamics of trust for each trial

long. However, this result except for one driver may suggest a nonlinear tendency between trust and driver's response to braking. In other words, drivers who hold appropriate, rather than excessive, trust in warning systems may avoid delaying their braking response when presented with a lack of alarms.

*Decreased of trust caused by two types of malfunctions*. Fig. 11 shows how the subjective rating of trust shifts during each trial. This figure indicates that a lack of alarms decreases trust more seriously than false alarms do. A two-way ANOVA on the subjective rating of trust shows a significant interaction of trials with types of malfunctions ($F(2,117)=17.32, p<0.01$). This result suggests that a decrease of trust caused by malfunctions may depend on malfunction type: false alarms and a lack of alarms.

## 5. DISCUSSION

This paper focused on how malfunctions in warning systems create a reduction in trust and therefore change the behavior of drivers.

The analysis of drivers' behavior showed that false alarms increased warning reaction time. The subjective ratings also showed that false alarms decreased trust. The above two results suggest that the decrease in trust due to false alarms gives cause to question reliance on warning systems. That is to say that owing to experience of false alarms, the warning system is not trusted by drivers, consequently the lack of trust in the system may lead to an increase in reaction time to warnings, even in functioning systems.

In addition to describing false alarms this article has considered a lack of alarms. The analysis of braking reaction time showed that a lack of alarms contributed to extending braking reaction time. Moreover breaking reaction time with the warning system equipped but non-operational was significantly longer than that without the warning system. The analysis of the relationship between the delay in reaction time and trust in the warning system showed that for drivers who hold a high level of trust in warning systems, their reaction time in response to emergency conditions may extend. Overall these results indicate that introducing a new driver support system, such as a rear-end collision warning system may lead to failures to act promptly to emergency conditions owing to dependence on the warning system, when alarms fail to operate.

It is difficult for designers to design completely reliable warning systems. The effectiveness of warning systems may be very high in correctly functioning situations, but when malfunctions occur, effectiveness cannot be assured. This research suggests that in order to maintain the maximum effectiveness of warning systems, it is essential for designers to cope with the effects of human trust in these systems deteriorating or rising excessively. Further investigation is needed to establish the appropriate level of trust that one can reasonably expect drivers to have in such systems to guarantee their efficiency.

The analysis of the influence of different kinds of malfunctions on trust showed that a lack of alarms resulted in a more significant distrust of warning systems than did false alarms. This evidence suggests that the way in which trust in waning systems decreases is a function of these types of malfunctions.

Trust may determine the usage of warning systems. Drivers may not use warning systems they believe to be untrustworthy. Therefore, the malfunction of warning systems may directly lead to disuse or then being ignored. In this respect, a lack of alarms may lead to dramatically decrease trust in warning systems. Therefore it may be necessary for designers to consider not only system reliability but also the type of malfunctions, and their relative frequency, which occur in the warning systems, so as to prevent drivers from dismissing warning systems.

## ACKNOWLEDGMENTS

The authors express special thanks to Dr. Tsuyoshi Katayama, Mr. Hiroshi Uno and Ms. Takako Mukasa, Japan Automotive Research Institute, Safety and IT Research Division, for their helpful comments of this study.

## REFERENCES

Itoh, M., G. Abe and K. Tanaka (1999). Trust in and use of automation: Their dependence on occurrence patterns of malfunctions. *Proc.IEEE SMC Conference*, 715-720.

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

Lee, J.D. and B. Kantowitz (1998). Perceptual and cognitive aspects of intelligent transportation systems. In: *Human Factors in Intelligent Transpotation Systems* (Barfield and Dingus (Ed)), LEA, 31-54.

Moray, N., T. Inagaki and M. Itoh (2000). Situation adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology Applied* , 6(1), 44-58.

Muir, B. (1987). Trust between human and machines, and the design of decision aids. *Int. J. Man-Machine Studies*, 27, 527-539.

Muir, B. (1994). Trust in automation: part I. Theoretical issues in the study of trust and human intervention in automation systems. *Ergonomics*, 37(11),1905-1922.

Muir, B. and N. Moray (1996). Trust in automation: part II. Theoretical issues in the study of trust and human intervention in automation systems. *Ergonomics*, 39(3), 429-460.

Parasuraman, R. and V. Riley (1997). Human and automation: Ues, misuse, disuse, abuse, *Human Factors*, 39(2), 230-253

Riley, V. (1996). Operator reliance on automation: Theory and data. In: *Automation and Human Performance* (Parasuraman and Mouloua(Ed)), LEA, 19-35.

Wilson, T.B., W. Butler, T.V. Maghee and T.A. Dingus (1997). Forward-looking collision warning system performance guideline. *SAE* Paper, 970456