Title

Toward overtrust-free advanced driver assistance systems

Author name
Makoto Itoh

Affiliation
Department of Risk Engineering,
University of Tsukuba

Address
1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573 Japan

email
itoh@risk.tsukuba.ac.jp

Tel
+81-29-853-5502

Fax
+81-29-853-5809

Abstract (150-250 words)

Avoiding human overtrust in machines is a vital issue to establish a socially acceptable Advanced Driver Assistance System (ADAS). However, research has not clarified the effective way of designing an ADAS to prevent driver overtrust in the system. It is necessary to develop a theoretical framework that is useful to understand how a human trust becomes excessive. This paper proposes a trust model by which overtrust can be clearly defined. It is shown that at least three types of overtrust are distinguished on the basis of the model. As an example, this paper discusses human overtrust in an Adaptive Cruise Control (ACC) system. By conducting an experiment on a medium-fidelity driving simulator, we observed two types of overtrust among the three. The first one is that some drivers relied on the ACC system beyond its limit of decelerating capability. The second one is that a driver relied on the ACC systems by expecting that it could decelerate against a stopped vehicle. It is estimated through data analysis how those kinds of overtrust emerged. Furthermore, possible ways for prevention of human overtrust in ADAS are discussed.

Keywords (4-6)
Overtrust, Advanced Driver Assistance Systems, Trust in automation,
Adaptive Cruise Control, Reliance, Human Interface

1 Introduction

Driver assistance systems are getting powerful and intelligent enough to control a vehicle. For example, Adaptive Cruise Control (ACC) systems (ISO, 2002) may perform longitudinal control to maintain the vehicle speed as directed by the driver if there is not a slow preceding vehicle, and to maintain the time gap to the preceding vehicle appropriately if there is one. ACC systems are useful to reduce driver mental workload so that the risk of the rear-end collision could be reduced as a result (see, e.g., Ma and Kaber, 2005; Young and Stanton, 2004). Also, automatic brake systems for collision avoidance have been studied (Coelingh, Eidehall, and Bengtsson, 2010; Isermann, Mannale, and Schmitt, 2010; Kaempchen, Schiele, and Dietmayer, 2009; Wada, Doi, Tsuru, Isaji, and Kaneko, 2010). Recently, automatic brake systems that work in the low-speed range have been putting into market (see, e.g., Distner, Bengtsson, Broberg, and Jakobsson (2009)).

However, the capability of a driver assistance system is essentially limited in some way. For example, a conventional ACC system does not apply the brake against stationary objects, which include a stopped vehicle at the tail end of a traffic jam. The main reason for ignoring stationary objects on the road is that a laser-radar in an ACC system for target vehicle detection can not distinguish reflectors on the guardrails from that on the tail of forward vehicles. In order to avoid unnecessary braking against reflectors on the guardrails, ACC systems should ignore reflectors not moving or whose relative speed with the host vehicle is very big. The speed threshold to ignore the preceding vehicle would depend on car models, and the exact threshold value is sometimes confidential. However, some carmakers disclose the threshold value via the owner's manual.

For example, one commercialized ACC system in Japan ignores a slowly moving object whose speed is below 20km/h. No systems can be free from this kind of capability limitation. Nilsson (1995) found in a simulator experiment that some drivers failed to intervene when approaching a stopped queue of vehicles because they believed that the ACC could effectively respond to the situation. Larsson (2011) found that users of an ACC system in the real world did not understand the limitation of the ACC system adequately. Avoiding driver overtrust in automation is thus important for attaining safety in the automobile domain.

Overtrust has been one of important issues in human factors (see, e.g., Parasuraman and Riley, 1997). It seems that overtrust in automation is closely related issues of "risk compensation" (Wilde, 1994) or "behavioral adaptation" (OECD, 1990). A driver may change his or her behavior into riskier one intentionally based on the recognition of risk reduction given by an automated system if the driver places his or her trust in the system very much (Itoh, Sakami, and Tanaka, 2007; Rudin-Brown and Parker, 2004). However, the driver behavioral adaptation depends on driving style (Hoedeaeker and Brookhuis, 1998). The notion of overtrust is also related to issues of "complacency." Even though discussions on defining complacency have not ended (see, e.g., Parasuraman, Sheridan, and Wickens, 2008), the term complacency refers to lack of vigilance (Moray and Inagaki, 2000) rather than intentional behavioral change towards compensation of the risk. According to Singh, Molloy and Parasuraman (1993), the tendency to be complacent is also dependent on the human attitude.

Even a vigilant human, on the other hand, may still trust an automation too much if he or she misunderstands what the automation can do (Itoh, 2010). In fact, Ockerman and Pritchett (2000) discussed professional workers' over-reliance on task guidance systems.

Unfortunately, it has not been established the way of avoiding driver overtrust in automation. It would be because that it has not been fully understood why human overtrust in automation occurs. Even though huge number of studies have been done for modeling of human trust in automation (see, e.g., Dzindolet, Pierce, Beck, and Dawe, 2002; Lee and Moray, 1992; Lee and See, 2004; Muir 1994; Sheridan 1992), the focus of many tends to be acceptance and promoting trust. The relationship between overtrust and trust has been hardly discussed. Although Lee and See (2004) discussed trust calibration, but the multidimensional aspects of trust were not taken into account in the paper. As has been pointed out, there exist multiple dimensions of trust in automation (Lee and Moray, 1992; Muir, 1994). The dimensions should be taken into account in the theory on overtrust in automation.

This paper proposes a model of human trust in automation which can give a clear definition of overtrust in automation. In order to develop design guidelines for prevention of overtrust, it is necessary to understand how human overtrust emerges. This paper conducts an experiment by using a driving simulator and analyzes the change of driver behavior.

2 Model of Trust and Overtrust

2.1 Trust and trustworthiness

[Definition 1: Trustworthiness of Automation]

Let *S* be a random variable representing a current situation in which an automated system is going to work. The random variable *S* can take on an element of $S_d = \{s_1, s_2, \cdots, s_{N_d}\}$ that is the set of all possible and mutually exclusive situations in which the automation is designed to work. The probability that the situation s_i occurs is represented as $P_d(S = s_i)$, where the subscript *d* represents the probability comes from the design of the system (Apparently, $\sum_{i=1}^{N_d} P_d(S = s_i) = 1$.). Event *A* represents that the automation carries out its required function, *Fa,* successfully. The objective trustworthiness of the automation under condition *S* can be expressed as the probability of *A* under situation *S*, i.e., $P_d(A \mid S)$. Thus, the overall trustworthiness of the automation, T_d, is defined as follows:

$$T_d = \sum_{i=1}^{N_d} P_d(A \mid S = s_i) P_d(S = s_i) \tag{1}$$

If the automation is designed and manufactured appropriately, the value of T_d should be one. However, T_d can be less than one due to unforeseen factors or events.

For example, let us think about the deceleration by an ACC system. Suppose the maximum deceleration rate of the ACC system is 0.25G, and every situation can be categorized into one of the following cases in terms of the necessary deceleration rate of the host vehicle:

s_1: 0.05 G is enough,
s_2: more than 0.05 G is necessary but 0.15 G is enough, and
s_3: more than 0.15 G is necessary but 0.25 G is enough.

Suppose the ACC system works perfectly in situations s_1 and s_2 (i.e., $P_d(A \mid S = s_1) = P_d(A \mid S = s_2) = 1$), but sometimes fails to decelerate at 0.25G at an s_3 situation. If the failure occurs once in a 10 occurrences of s_3 and the probabilities of occurrence of s_1, s_2, and s_3 are 0.5, 0.3, and 0.2, respectively, the total trustworthiness of the ACC system T_d is obtained as follows:

$$T_d = 1 \times 0.5 + 1 \times 0.3 + 0.9 \times 0.2 = 0.98$$

It is true that there exist situations where more than 0.25G is necessary as the deceleration rate to avoid a collision, but those situations are out of the scope of the system. Therefore, the incapability of the ACC system beyond the brake limit does not affect the trustworthiness of the system.

[Definition 2: Human Operator's Trust in Automation]

Let *S'* be a random variable representing a current situation in which the human operator expects that the automated system is going to work. The random variable *S'* can take on an element of $S_h = \left\{s'_1, s'_2, \cdots, s'_{N_h}\right\}$ that is the set of all possible and mutually exclusive situations in which the human operator regards the automation may work well. The probability that the situation s'_i occurs is represented as $P_h(S' = s'_i)$, where the subscript *h* represents the probability is estimated by the human operator, and again, $\sum_{i=1}^{N_h} P_h(S = s_i') = 1.$). Suppose the human operator's degree of belief on trustworthiness of the automation under situation *S'* is expressed as a subjective probability denoted by $P_h("A" | S')$, where "*A*" represents that the human operator expects the automation carries out its expected function, F_h, under situation *S'* successfully. The human operator's overall trust in the automation, T_h, is defined as follows:

$$T_h = \sum_{i=1}^{N_h} P_h("A" | S' = s'_i) P_h(S' = s'_i) \tag{2}$$

Note here that $S_d \neq S_h$ in general.

The above definitions of the automation trustworthiness and the human trust in the automation imply that the 100% trust does not always mean overtrust. If the trustworthiness is perfect, it would be appropriate that a human operator trusts the automation completely. Such trust is never excessive.

Again, let us think about the deceleration by the ACC system. Suppose a driver regards that the maximum deceleration rate of the ACC system is approximately 0.2G, and categorizes situations as follows:

s'_1: 0.1 G is enough, and

s'_2: more than 0.1G is necessary but 0.2 G is enough.

Suppose the driver has experienced up to 0.2G deceleration of the ACC system, and the system has worked perfectly so that the driver has been satisfied with the ACC system (i.e., $P_h("A" | S' = s'_1) = P_h("A" | S' = s'_2) = 1$). If the subjective probabilities of s'_1 and s'_2 are 0.6 and 0.4, respectively, the overall f trust T_h is obtained as follows:

$T_h = 1 \times 0.6 + 1 \times 0.4 = 1.0$

Apparently, this should not be regarded as overtrust.

2.2 Overtrust

How can human operator overtrust in automation be defined? It could be regarded as overtrust if $T_h > T_d$ in some cases. However, as shown in the examples in section 2.1, the fact $T_h > T_d$ itself does not always mean overtrust. Moreover, it is difficult to develop useful countermeasures even if we notice the occurrence of overtrust by knowing $T_h > T_d$. Based on the definitions 1 and 2, this paper regards the human operator's trust as excessive if at least one of the following conditions is satisfied.

(i) $P_h(\text{"}A\text{"} \mid S' = s) > P_d(A \mid S = s)$, where $s \in S_d$
(ii) $S_h - S_d \neq \phi$ (null set)
(iii) F_h is not equivalent to F_a.

Note here that the case (i) should be impossible if $s \in S_d$, because the value of $P_d(A \mid S = s)$ should be one in that situation in principle. In reality, however, it would be possible that $P_d(A \mid S = s)$ is less than one even if $s \in S_d$. The overtrust in this sense is related to the lack or reduced vigilance against system malfunctioning. However, the principal task here is to improve the system reliability.

Case (ii) means that the human operator expects the automation carries out its function beyond the situations specified in the system design. A typical example is that a driver expects that an ACC system can prevent a rear-end crash when the necessary deceleration is higher than the maximum deceleration rate of the ACC system. The overtrust here is related to the "performance" dimension of trust (Lee and Moray, 1992). This type of overtrust has been observed in several studies (see, e.g., Itoh, 2007; Seppelt and Lee, 2007). It is hypothesized that this type of overtrust is caused by driver's direct extrapolation of subjective expectation of system capability from the previous experience to the non-experienced situations. That is, the increase of $P_h(\text{"}A\text{"}|s'_i)$ results in the increase of $P_h(\text{"}A\text{"}|\text{'}s_j)$, where s'_i is in S_h but s'_j is not. We call this expansion of driver expectation a ripple effect. Such ripple effects were observed in a very simple experiment in a process control system (Itoh, Inahashi, and Tanaka, 2003).

However, it is not clear whether such ripple effects occur in a realistic situation of car driving.

Case (iii) represents that the human misunderstands the function of the automation and he/she expects the work that the automation is not designed to provide. This is related to "purpose" dimension of trust (Lee & Moray, 1992). A typical example of such overtrust is driver expectation to an ACC system to decelerate against a stopped vehicle at the tail end of a traffic jam. Dickie and Boyle (2009) showed that many drivers were not familiar with this limitation. What happens if a driver comes to a stopped vehicle at the tail end of a traffic jam? Does the driver overtrust appear? If so, what are the contributing factors for the overtrust? It is necessary to conduct an experiment to find answers to the questions.

3 Experiment

3.1 Purpose

The purpose of the experiment is to test whether driver overtrust in an ACC system can be observed in driving simulator experimental conditions. If yes, it is also necessary to clarify reasons why such overtrust occurs, in order to establish methodologies for prevention of overtrust.

3.2 Method

3.2.1 Participants

Twelve drivers (six females and six males) between the ages of 26-55 years (mean = 37.0, s.d.=8.6) participated in this experiment. Every participant had a valid driver’s license and drove daily. The driving experience was more than 10 years for every participant except one who had been a licensed driver for less than one year.

3.2.2 Apparatus

A fixed-base driving simulator was used in this study (Fig. 1). It has a nearly straight, two-lane, and endless expressway. Three 100-inch screens are set in front of the driver and the field of view is approximately 120 degrees. An

ACC system is installed in this driving simulator. The main characteristics of the ACC system are as follows:

- The ACC system controls the vehicle speed at the target level set by the driver when there is no lead vehicle ahead. The system maintains the safe headway distance when a lead vehicle exists.
- The ACC system can be activated and work from 5km/h to 100km/h. The driver has to apply the brake for the full stop.
- The ACC system is activated only if the driver presses the activate button near the steering wheel, but can be deactivated by pushing the cancel button or by pressing the brake pedal.
- A visual icon appears on the control panel including the speed meter while the system recognizes the lead vehicle. The icon disappears if the lead vehicle is lost. It is assumed that no error occurs in the system for the detection of a lead vehicle and the estimation of the headway distance and the relative speed.
- The maximum system acceleration rate is 0.15 G, and the maximum system deceleration rate is 0.25 G. Neither auditory nor visual information is given even when the ACC system is conducting the maximum brake (Note: This is not the standard configuration of ACC systems. In the real world, ACC systems may provide some information to driver at the maximum deceleration. The reason for not issuing an alert in this experiment was to avoid driver simple reaction to the alert.).
- The ACC system can not detect stationary objects. No assistance is given against the stationary objects on the road.

****

Insert Figure 1 about here

****

3.2.3 Task, experimental design, and procedure

The participants were instructed to drive safely on the left-hand lane of the expressway by using the ACC system as much as possible. At the end of a drive, the host vehicle comes to the tail end of a heavy traffic jam. All vehicles in front are stopped. When the host vehicle fully stops, the trial ends.

In terms of the situation before arriving at the tail end of the traffic jam, six scenarios (five 5-min drive scenarios from A1 to A5 and one 20-min drive scenario named B) were distinguished:

A1: The host vehicle was running at a cruise speed and came close to the slow lead vehicle running at 50 km/h. After following the lead vehicle at 50km/h for a while, the host vehicle loses the lead vehicle because the lead vehicle changes lanes just before arriving at the tail end of the traffic jam. The ACC system in the host vehicle is still active even after losing the lead vehicle, but the system does not provide any help against the stopped vehicles ahead. The driver in the host vehicle should apply the brake by him/herself for rear-end collision avoidance.

A2: After free cruising, the host vehicle came close to a vehicle whose speed was 50km/h and made following for a while. Then the lead vehicle decelerates at 0.1G and finally stops because of the traffic jam. The ACC system in the host vehicle decreases the vehicle speed accordingly and successfully. At least, the driver in the host vehicle has to apply the brake for stopping when the vehicle speed becomes below 5km/h.

A3: The host vehicle was following the lead vehicle at 100km/h. The lead vehicle decelerates at 0.1G upon detecting the traffic jam and finally stops. The ACC system in the host vehicle decreases the vehicle speed accordingly and successfully. At least, the driver in the host vehicle has to apply the brake for the stopping.

A4: The host vehicle was following the lead vehicle at 100km/h. The lead vehicle decelerates at 0.2G upon detecting the traffic jam and finally stops.

The ACC system may apply almost its maximum brake against the lead vehicle deceleration, but anyway it is successful. Again, the driver in the host vehicle has to apply the brake for stopping.

A5: The host vehicle was following the lead vehicle at 100km/h. The lead vehicle decelerates at 0.35G upon detecting the traffic jam and finally stops. Even though the ACC system applies its maximum brake, a rear-end crash can not be avoided only by the system brake. The driver has to intervene into the control as soon as possible.

B: The host vehicle was following the lead vehicle at 100km/h. Before arriving at the tail end of the traffic jam, the lead vehicle changes lanes. The ACC system in the host vehicle is still active even after losing the lead vehicle, but the system does not provide any help against the stopped vehicles ahead. The driver in the host vehicle should apply the brake by him/herself for rear-end collision avoidance.

The ACC system works perfectly as it is designed, no malfunction occurs in the system. However, the participants did not receive any information on the reliability of the ACC system at all. The participants were informed that there is a limitation of the deceleration rate given by the ACC system but not informed the actual value of the limitation. No information was given to the participants on the system behavior against stopped objects.

Every participant received all six scenarios in the data collection. A trial has one scenario. The number of trials for each scenario and the order of scenarios are shown in Table 1. The numbers of the scenarios are not balanced: ordinary decelerations of the lead vehicle, such as 0.1G, are many, but the rapid decelerations, such as 0.35G, are few.

****

Insert Table 1 about here

****

The experiment lasted four days for each participant. In each day, it took approximately one hour depending on the number of trials to be completed. On the first day, participants were given opportunities for practice so that they could become familiar with the simulator and with the ACC system. There were neither rapid decelerations of the lead vehicle nor stopped objects in the practice drives. The practice drives were done until the driver felt that he or she had obtained enough skill for driving on the simulator. In the data collection, a short break was given every 6 or 7 trials for every participant depending on the driver state. Before conducting scenario B on the fourth day, every driver was given a short break. The participants were informed of neither the number of trials in a day nor the content in each scenario so as to avoid the driver's unnecessary prediction of events in the trials. We were concerned that participants might "predict" something wrong may happen at the final trial on the last day.

3.2.4 Dependent variables

In this paper, one of the most important dependent variables is the number of rear-end collisions.

In order to investigate how driver attitude toward the ACC system changed through his or her repetitive use of it, we analyze driver brake timing in type A scenarios from the first day to the last day. An appropriate index for discussing the brake timing depends on scenarios. For scenarios A1 and B, Time To Collision (TTC) against the stopped vehicle at the tail end of the traffic jam would be an appropriate one. For scenarios A2, A3, A4, and A5, TTC against the lead vehicle to follow may not be appropriate, because the value of TTC becomes huge if a driver waits until the vehicle speed becomes below 5km/h (minimum working speed of the ACC system). Instead, the host vehicle speed at the driver brake would be the appropriate one. If the vehicle speed at the braking is large (small), the brake timing can be regarded as early (late).

3.3 Results and discussions

No crashes occurred in the total 24 cases (2 trials/participant * 12 participants) of scenario A1 trials. Among 84 cases of A5 in which the lead vehicle decelerated rapidly, three crashes occurred. Participants #1, #4, and #12 caused a crash for each. For scenario B, there was one crash caused by participant #11 among 12 cases. According to surveillance of video images of the driver face at the crashes, the four crash-experienced participants were not drowsy, not distracted, but looked forward. Fig. 2 shows the values of 1/THW and 1/TTC at the driver brake onset. Goodrich and Boer (2003) suggest that driver risk perception on a rear-end collision can be described with the combination of THW and TTC. On the basis of that, Kondoh, Yamamura, Kitazaki, Kuge, and Boer (2008) investigated real world driver braking behavior and showed that the value of RF=1/THW+4/TTC tends to be lower than two at the driver brake onset in ordinary situations (Note: RF represents “Risk Feeling.”). The data shown in Fig. 2 suggest that the driver brake in the crash cases was late.

****

Insert Figure 2 about here

****

Thus, it can be claimed that the reliance of the crash-experienced participants on the ACC system was too high at the crash. For the crash cases in type A5, their reliance on the ACC system was excessive in the sense that they let the ACC system brake when the necessary deceleration rate was beyond the limit of the system capability. On the other hand, the reliance on the system was excessive in the sense that the crashed participant misunderstood the purpose of the system. According to the comments given at the interview after the experiment, the participant expected the ACC system to reduce the vehicle speed to some extent at that situation, but in fact, the system never applies the brake against the stopped vehicle.

Why did the crashes occur? How we can reduce the risk of such crashes? As has been shown in Fig. 2 (a), the lack of information on the boundary of the system capability does not always cause a crash. Other contributing factor(s) is/are necessary for occurrences of crashes. This paper focuses on effects of successful use of the ACC system under peaceful situations, i.e., we investigate how driver behavior changed on the basis of driving experience with the ACC system. Concretely speaking, how the timing of driver brake, which means driver takeover of control from the ACC system, is changed on the basis of repetitive use of the ACC system. Fig. 3 shows the time series of driver brake timing for each scenario type. Fig. 3 suggests the followings:

(1) Scenario A3. The crash-experienced participants, #1, #4, #12, and #11, were very reliant on the ACC system. Among non-crash experienced participants, there were two participants, #5 and #7, who were not willing to use the ACC system. It seems that the two participants developed their rule to intervene into control; they applied the brake immediately upon detecting deceleration of the preceding vehicle instead of letting the system brake. For the remaining participants, the brake timing in trials of scenario A3 became late gradually. The ACC system was worth relying for those participants except #5 and #7 in that situation.

(2) Scenario A5. Experience of a crash or a near-miss at a trial made brake timing earlier at least for the next several trials, where a near-miss refers to a case in which RF > 2 at the driver braking. The experience of a crash or a near-miss makes a driver attentive.

(3) Scenario A4. The experience of a crash or a near-miss in a type A5 trial makes the driver brake timing earlier in the type A4 trial just after the crash or the near miss.

****

Insert Figure 3 about here

****

There was a tendency that the brake timing at scenario B was early for the participants who were not willing to use the ACC system (#5 and #7) and who experienced a crash in a trial of A5 scenario (#1, #4, and #12), compared to the other participants (Fig. 4). The reason why the brake timing of participants #1,

#4, and #12 was early would be that their attention was aroused by the experience of the crash in a type A5 trial.

****

Insert Figure 4 about here

****

In summary, it would be possible to categorize the participants as shown in Fig. 5. The questions here are: (1) What are the differences between the three crash-experienced participants in a trial of scenario A5 (#1, #4, and #12) and the other 10 participants who were willing to use the ACC system, and (2) What are the differences between participant #11 and the remaining six?

****

Insert Figure 5 about here

****

In order to answer to the first question, the following analyses were done. A t-test was conducted to compare the brake timing in A3 trials between the three crash-experienced participants and the remaining 7 participants. The result showed that there was a statistically significant difference between them ($t(442)=-4.6$, $p<0.01$). A t-test on the brake timing in A4 trials between the three crash-experienced and the seven non-experienced showed a significant difference between them ($t(212)=-3.3$, $p<0.01$). The crashes in A5 trials occurred due to the delay in driver braking.

Thus, the reliance on the ACC system can be illustrated as shown in Fig. 6(a). Fig. 6(b) depicts a quantitative estimation of the reliance. In Fig. 6(b), a point represents the mean value of the estimated willingness levels to rely on the ACC system in the corresponding participant group (crash-experienced: #1, #4, #11, and 12, not crash-experienced: others), and the error bar represents the standard deviation. The estimated willingness level (WL) is derived as WL = 1 - (the vehicle speed at braking) /100, where the unit of the vehicle speed at braking is km/h, and the value is obtained from the final trial of the corresponding scenario (A3: Day #4, Trial #13, A4: Day #4, Trial #7, and A5: Day #4, Trial #11).

Moreover, the correlation between a brake timing at an A5 trial and the brake timing at the preceding A4 trial of the A5 trial was 0.54 ($p<0.05$) (Fig. 7). On the basis of the above results, it could be claimed that an experience at a condition may affect the willingness to rely on the ACC system at another condition as shown in Fig. 8. Concretely speaking, the experience of the success of ACC increases not only the reliance on the system at A4 but also the reliance on the system at A5. This effect was stronger for the crash-experienced participants than for the other participants. Note here that the ripple effect model shown in Fig. 8 has not been verified in a quantitative manner. Further research is necessary to develop a model which is able to describe the dynamics of the trust.

****

Insert Figure 6 about here

****

****

Insert Figure 7 about here

****

****

Insert Figure 8 about here

****

As for the second question, participant #11 had relied on the ACC system until it became deactivated in almost all A3 trials (Fig. 3(a)), but the remaining six had not (Fig. 3(b)). The brake timing of participant #11 in A5 became late gradually (Fig. 3(e)), but that of the other six did not. This increase of participant#11’s reliance on the ACC system in scenario A5 may have a strong relationship with the reliance on the system in scenario B.

Participant #11 was closer to #1, #4, and #12 than to the other six drivers. It was common for the four participants in the sense that they became reliant on the ACC system too much in A5 trials. The difference between participant #11 and participants #1, #4, and #12 was that the former experienced neither a crash nor a near-miss in A5 trials due to the gradual increase of the reliance. It is thus possible that the participants #1, #4, and #12 caused a crash in scenario B if they had faced scenario B before they experienced the crash in an A5 trial.

4 Conclusions

This paper proposed a model of human trust in automation for discussing overtrust. It is necessary to distinguish at least overtrust in terms of performance from overtrust in terms of purpose.

The results of the experiment suggest that increase of trust in a system at some working condition may cause a ripple effect to trust in it in more difficult working conditions. The ripple effect seems very natural but has hardly been observed in previous experimental studies on human-machine cooperation. Overtrust due to misunderstanding of the system purpose was also observed in this experiment. According to the investigation of driver behavior, the false expectation towards the ACC system to decelerate against a "stopped" lead vehicle may be strongly related to the repetitive driver observations of successful system behavior against a "stopping" lead vehicle.

How can we apply the observations to system design for prevention of human overtrust in automation? Essentially, the ripple effect would be inevitable in the process of development of human trust in automation, because trust in another person often emerges like that. For example, a human supervisor asks a human subordinate to do an easy task. After observing that the subordinate completes the task successfully, the supervisor may expect that the subordinate will be able to do a more difficult task. The ripple effect is adequate in a human-human relationship because the skill of a subordinate may be developed gradually. It may not be appropriate, on the other hand, for a human to expand the subjective limit of automation capability on the basis of the experience, because an automated system does not grow as a human does. An automated system should provide information on the purpose of the system and the limit of its capability in a clear manner.

Stanton and Young (2005) suggested the necessity of visual information for driver prediction of system behavior. Seppelt and Lee (2007) proposed an ecological interface display which makes the limits of an ACC system visible to the driver. Itoh (2008) proposed a method of displaying the limit of deceleration capability of an ACC system on the instrumental panel for preventing overtrust in terms of performance. However, the methodologies for prevention of overtrust due to misunderstanding of the purpose have not been established. Further studies are necessary to clarify whether overtrust in terms of purpose can be prevented by an appropriate design of human interface or human-machine interaction, or whether only education or training is the practical solution.

Note that it is still unclear whether or not overtrust observed in this paper really occurs in the real world. This is because a part of the configuration of the ACC system in this study is different from the real one. For example, the ACC system used in this study does not give any alert at the maximum braking. Further studies are necessary to observe driver behavior in the real world.

In this paper, the theoretical model of trust in automation proposed in section 2 was used only for discussing what kinds of overtrust are possible. However, the model itself has a potential to describe how the trust emerges. Further studies are necessary to investigate whether or not the overall trust in automation is really derived by formula (2). Another problem rises here. It might be possible that a human operator is not able to provide a subjective probability which satisfies the probability axioms. One possibility to overcome this problem is to apply the Dempster-Shafer theory of evidence (Shafer, 1976) which is suitable for describing human subjective feeling to which the probability theory is difficult to apply (Itoh, 2001).

Quantitative formalization of the ripple effect is also necessary. One way for this is to apply the model of dynamic change of trust proposed by Gao, Lee, and Zhang (2006). However, the application of the model of Gao et al. is not straightforward. The formalized quantitative modeling of dynamics of trust will be discussed in a different paper.

Acknowledgment

The author would thank Junko Kunugi for her help in conducting the experiment and in the data analysis.

References

Coelingh, E., Eidehall, A., & Bengtsson, M. (2010). Collision warning with full auto brake and pedestrian detection –a practical example of automatic emergency braking-, Proceedings of 2010 13th International IEEE Annual Conference on Intelligent Transportation Systems, Madeira Island, Portugal, 155-160. doi: 10.1109/ITSC.2010.5625077

Dickie, D. A., and Boyle, L. N. (2009). Driver's understanding of adaptive cruise control limitations, Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting, 1806-1810, San Antonio.

Distner, M., Bengtsson, M., Broberg, T., Jakobsson, L. (2009). City safety - a system adressing rear-end collisions at low speeds, Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles, No. 09-0371, Stuttgart, Germany. Retrieved from http://www-nrd.nhtsa.dot.gov/pdf/esv/esv21/09-0371.pdf

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task, Human Factors, 44(1), 79-94.

Gao, J., Lee, J.D., Zhang, Y. (2006). A dynamic model of interaction between reliance on automation and cooperation in multi-operator multi-automation situations, International Journal of Industrial Ergonomics, 36, 511-526.

Goodrich, M.A., & Boer, E.R. (2003). Model-based human-centered task allocation: a case study in ACC system design, IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 33(3). 325-336.

Hoedemaeker, M., & Brookhuis, K.A. (1998). Behavioral adaptation to driving with an adaptive cruise control (ACC), Transportation Research Part F: Traffic Psychology and Behaviour, 1(2), 95-106.

Isermann, R., Mannale, R., & Schimitt, K. (2010). Collision avoidance systems PRORETA: situation analysis and intervention control, Proceedings of 6th IFAC Symposium in Advances in automatic Control, 461-470, Munich, Germany.

ISO (2002). Transport information and control systems -- Adaptive Cruise Control Systems -- Performance requirements and test procedures, 15622.

Itoh, M. (2001). Ambiguity as a buffer in change of trust in automation, Proc. IFAC Symposium in Analysis, Design, and Evaluation of Human-Machine Systems, Kassel, Germany, 241-247.

Itoh, M., Inahashi, H., Tanaka,K. (2003). Informing Limit of Ability and its Reasons for Reducing Overtrust in Automation, Proc. XVth Triennial Congress of the International Ergonomics Association, Seoul, 1: 141-144.

Itoh, M., Sakami, D., & Tanaka, K. (2007). Risk Compensation due to Human Adaptation to Automation for System Safety, Transactions of the Society of Instrument and Control Engineers, 43(10), pp. 926-934.

Itoh (2008). Deceleration Meter: A Management Tool for Reducing Over-reliance in Collision Warning when using Adaptive Cruise Control System, Proc. PSAM9, CD-ROM, 7 pages.

Itoh, M. (2010). Necessity of Supporting Situation Awareness to Prevent Over-trust in Automation, International Electric Journal of Nuclear Safety and Simulation, 1(2), 150-157.

Kaempchen, N., Schiele, B., & Dietmayer, K. (2009). Situation assessment of an autonomous emergency brake for arbitrary vehicle-to-vehicle collision scenarios, IEEE Transactions on Intelligent Transportation Systems, 10(4), 678-687.

Kondoh, T., Yamamura, T., Kitazaki, S., Kuge, N., & Boer, E. R.: Identification of Visual Cues and Quantification of Drivers' Perception of Proximity Risk to the Lead Vehicle in Car-Following Situations, Journal of Mechanical Systems for Transportation and Logistics, 1(2), 170-180.

Larsson, A.F.L. (in press). Driver usage and understanding of adaptive cruise control, Applied Ergonomics (Available on line from August 25, 2011)

Lee, J. D. and Moray, N. (1992). Trust, Control Strategies and Allocation of Function in Human-Machine Systems, Ergonomics, 3(10): 1243-1270.

Lee, J. D., and See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance, Human Factors, 46(1): 50-80.

Ma, Ruiqi, & Kaber, D.B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone, International Journal of Industrial Ergonomics, 35(10), 939-953.

Moray, N., and Inagaki, T. (2000). Attention and Cmplacency, Theoretical Issues in Ergonomics Science, 1(4): 354-365.

Muir, B. (1994). Trust in Automation: Part I. Theoretical Issues in the Study of Trust and Human Interaction in Automated Systems, Ergonomics, 37(11): 1905-1922.

Nilsson, L. (1995). Safety effects of adaptive cruise controls in critical traffic situations. Steps Forward Proceedings of the Second World Congress on Intelligent Transport Systems 95 Yokohama (Vol. 3, pp. 1254 - 9).

Ockerman, J., Pritchett, A. (2000). Reducing Overreliance on Task-Guidance Systems, Proceedings of the IEA2000/HFES2000 Congress, San Diego, 1: 1-4.

OECD (1990). Behavioral Adaptations to changes in the road transportation system. Paris: OECD.

Parasuraman, R., & Riley, V. (1997). Human and Automation: Use, Misuse, and Abuse, Human Factors, 39(2): 230–253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Construct, Journal of Cognitive Engineering and Decision Making, 2(2): 140-160.

Rudin-Brown, C.M., & Parker, H.A. (2004). Behavioral adaptation to adaptive cruise control (ACC): implications for preventive strategies, Transportation Research Part F, 7: 59-76.

Seppelt, B.D., & Lee, J.D. (2007). Making adaptive cruise control (ACC) limits visible, International Journal of Human-Computer Studies, 65, 192-205.

Shafer, G. (1976). A Mathematical Theory of Evidence, Princeton University Press, Princeton.

Sheridan, T. B. (1992). Telerobotics, automation, and human supervisory control, Cambridge, MA: MIT Press.

Singh, I. L., Molloy, R., Parasuraman, P. (1993). Automation-induced "Complacency": Development of the Complacency-potential Rating Scale, The International Journal of Aviation Psychology, 3(2): 111-122.

Wada, T., Doi, S., Tsuru, N., Isaji, K., & Kaneko, H. (2010). Characterization of expert drivers' last-second braking and its application to a collision avoidance system, IEEE Transactions on Intelligent Transportation System, 11(2), 413-422.

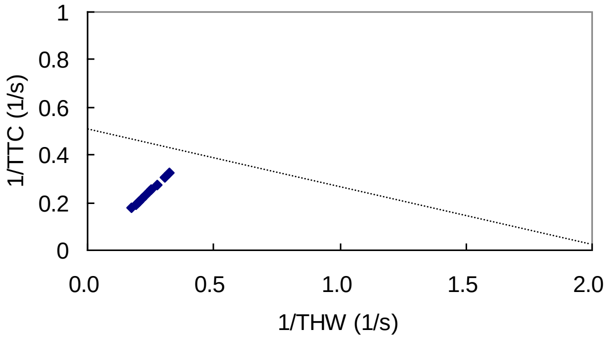
Wilde, G. J. S (1994). Target Risk, Toronto, Ontario, Canada: PDE Publications.

Young, M.S., Stanton, N.A. (2004). Taking the load off: investigations of how adaptive cruise control affects mental workload, Ergonomics, 47(9), 1014-1035.

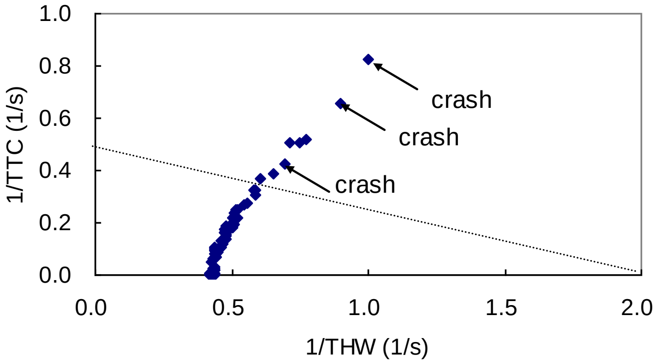
Table 1 Experiencing order of scenarios

Trial #	Day #1	Day #2	Day #3	Day #4
1	A3	A3	A3	A3
2	A4	A3	A4	A4
3	A3	A4	A3	A3
4	A3	A3	A2	A4
5	A4	A2	A3	A3
6	A3	A4	A4	A2
7	A4	A3	A3	A4
8	A3	A3	A5	A1
9	A5	A4	A3	A3
10	A3	A2	A4	A3
11	A4	A5	A3	A5
12	A3	A3	A1	A3
13	A4	A4	A3	A3
14	A3	A3	A3	B
15	A5	A4	A2	
16	A3	A3	A4	
17	A3	A3	A3	
18		A3	A5	
19		A5	A3	
20			A4	

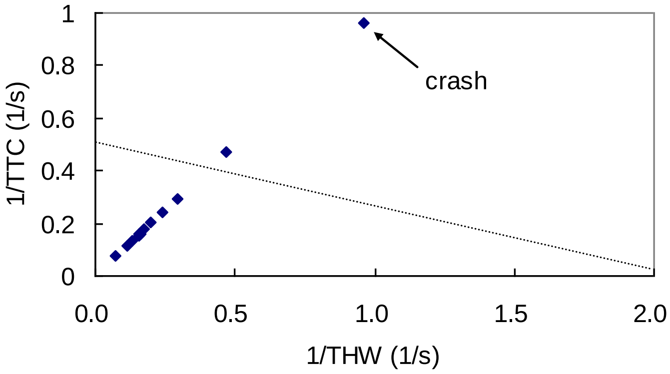
Fig. 1 The driving simulator used in the experiment



(a) Scenario A1

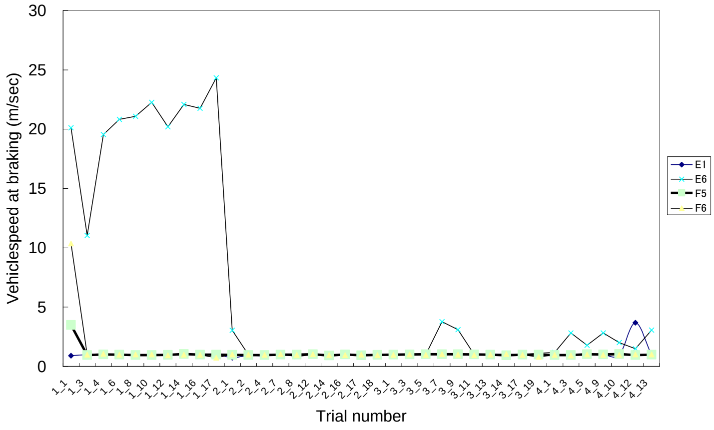


(b) Scenario A5

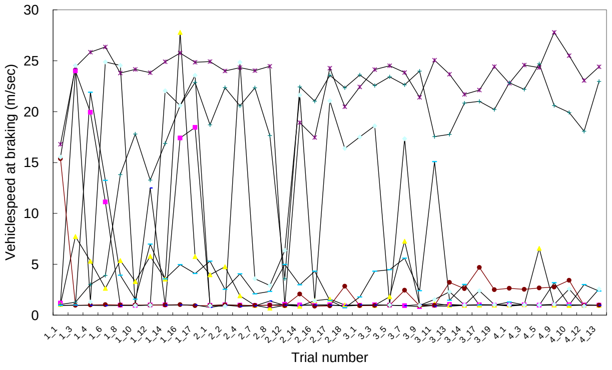


(c) Scenario B

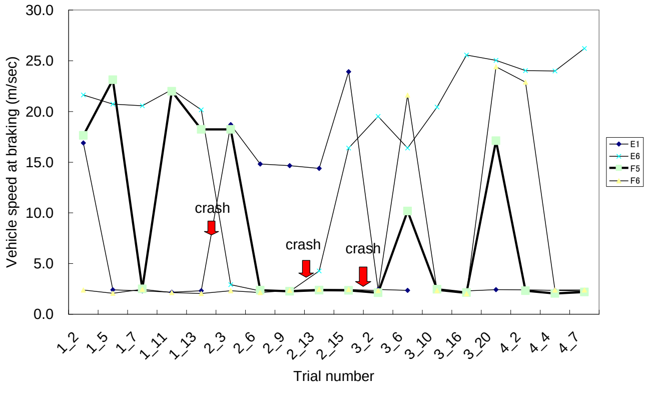
Figure 2 Driver brake timing in risky situations



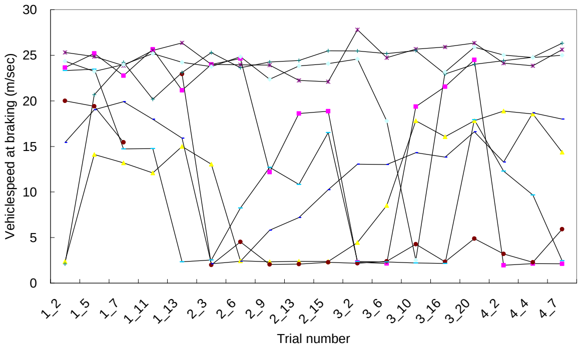
(a) scenario A3 (crash-experienced)



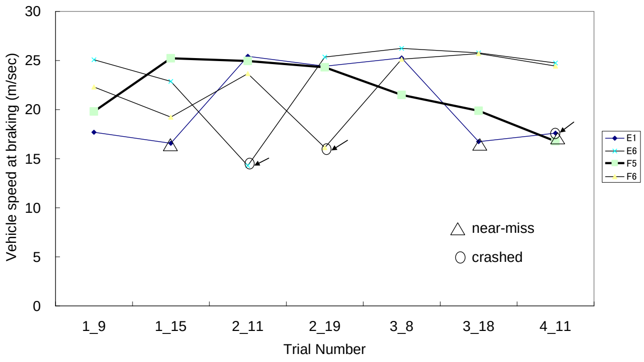
(b) scenario A3 (not crash-experienced)



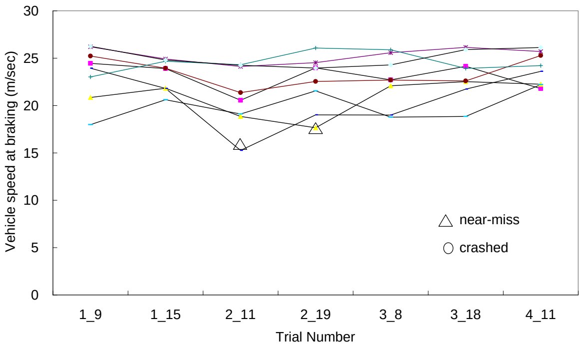
(c) scenario A4 (crash-experienced)



(d) scenario A4 (not crash-experienced)



(e) scenario A5 (crash-experienced)



(f) scenario A5 (not crash-experienced)

Figure 3 Driver brake timing for each scenario

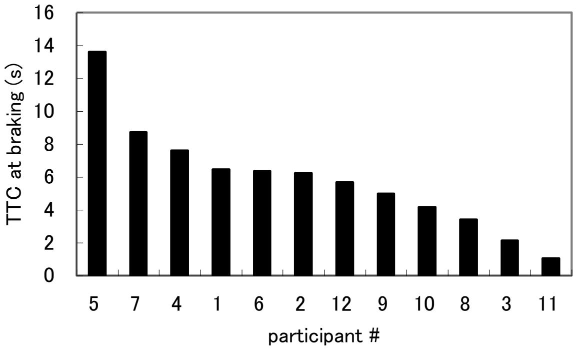


Fig. 4 TTC at driver braking in scenario B

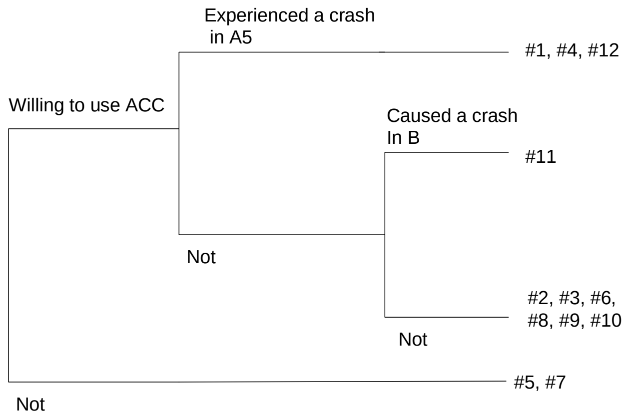
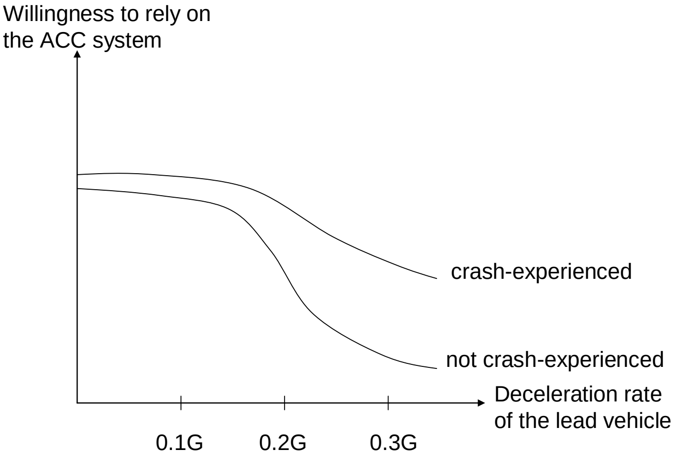
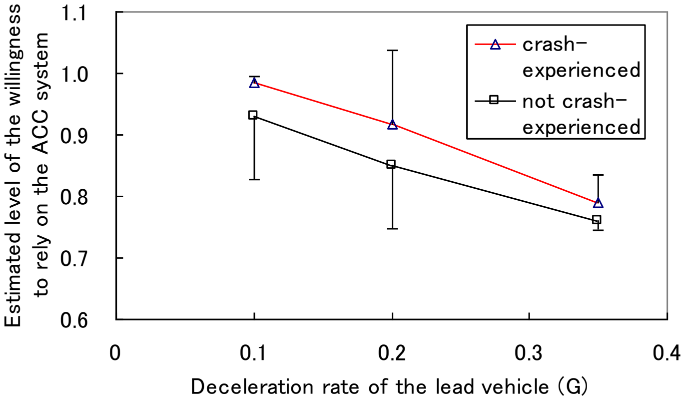


Fig. 5 Categorization of participants



(a) conceptual model



(b) estimated model

Fig. 6 Difference in the reliance on the ACC system between #1, #4, and #12 (experienced a crash in a A5 trial) and the others

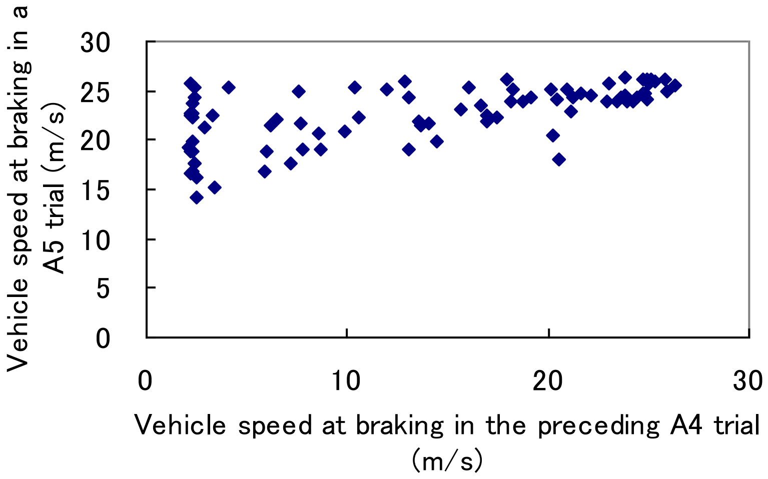


Fig. 7 Relationship between brake timings in an A5 trial and in its preceding A4 trial

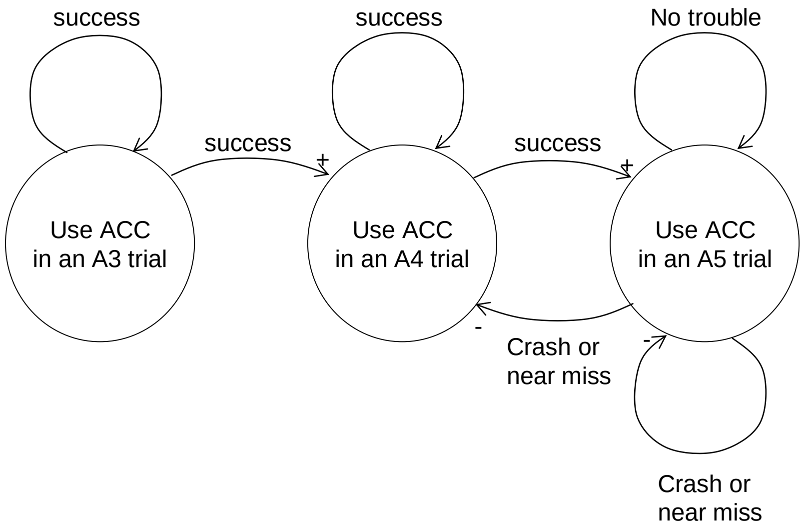


Fig. 8 The ripple effect