How Does Explanation-Based Knowledge Influence Driver Take-Over in Conditional Driving Automation?

Huiping Zhou, Makoto Itoh, *Member, IEEE*, and Satoshi Kitazaki, *Member, IEEE*

***Abstract*—This article focuses on explanation-based knowledge about system limitations (SLs) under conditional driving automation (society of automotive engineers level 3) and aims to reveal how this knowledge influences driver intervention. By illustrating the relationships between the driving environment, system, and mental model, knowledge in dynamic decision-making processing for responding to an issued request to intervene (RtI), occurrence of SL, concept of RtI, and scene(s) related to SL are determined by knowledge-based learning. Based on three concepts, the knowledge is examined at five levels: 1) no explanation, 2) occurrence of SL, 3) concept of RtI, 4) some typical scenes related to SL, and 5) all of the above. Data collection is conducted on a driving simulator, and 100 people with no experience of automated driving participated. The experimental results show that instructing drivers in typical situations contributes to a greater increase in the rate of successful intervention in car control from 55% to 95%. Furthermore, instructing them on the concept of RtI is conducive to a significant reduction in response time from 5.48 to 3.62 s in their first experience of RtI. It is also revealed that the knowledge-based learning effect dwindles but does not vanish even after drivers experience RtI a number of times. Compared to explaining all possible situations to a driver, introducing typical situations results in better take-over performances even in critical or unexplained scenarios. This article demonstrates the importance and necessity of this knowledge, especially the explanation of sample scenes related to SL, which contributes to drivers' take-over behavior.**

***Index Terms*—Conditional driving automation, driver takeover, human-machine interface, information processing, safety, system limitations.**

I. Introduction

The concept of automated driving has become widespread over the last decade as the capabilities of automated driving systems have rapidly increased, thereby raising the potential of such systems to not only improve traffic safety [1]–[3] but also ease drivers' mental workloads.

The development of automated driving systems has been a step-by-step process. According to the taxonomy of automated driving systems defined by the society of automotive engineers (SAE) in their J3016 [4] directive, the first achievement was Level 1 "driver assistance" systems. During the development of such systems, issues that influenced driving safety while the driver retained primary control, such as automation surprise [5], inattention and distraction [6], [7], excessive trust, and overreliance [8], were discussed and investigated.

Issues such as system limitation (SL) [9], automation complacency [10], and driver-vehicle interfaces [11], [12] under Level 2 "partial driving automation" conditions became serious because the driver could actually be physically removed from the decision-making process. Partial automation supplies more advanced control in comparison to the driving assistant system in terms of reliance on the system, but drivers still need to be constantly involved in its supervision.

Since users under Level 3, which is defined as "conditional automated driving" (see SAE J3016), are not required to constantly monitor the system and/or the surrounding environment while the automated driving system is operating normally, they have the option of engaging in additional nondriving-related secondary activities, such as watching videos or reading books, and it is much more likely that they will become fatigued [13]–[15]. However, they are still expected to smoothly intervene and resume control when the system requests them to take over car control, similar to the lower levels of automated systems.

In their study, Zeeb *et al.* [16] pointed out that drivers' readiness states influenced their take-over behavior when it became necessary to respond to an emergency situation and suggested that their take-over behavior would increasingly worsen according to the extent to which a driver was distracted by a nondriving-related task. Therefore, it is vital to monitor the driver's state to ascertain whether they are in a normal state of readiness to resume full control of the vehicle once the automated system issues a request to intervene (RtI).

In response to these concerns, numerous efforts have been made to assess driver readiness via gaze behavior [16], driving data, physical state (such as hand, feet, and seating position), and various combinations of such factors [17]. In addition, numerous studies have been conducted to determine the optimal

Manuscript received August 4, 2019; revised July 11, 2020 and December 16, 2020; accepted December 28, 2020. Date of publication February 9, 2021; date of current version May 19, 2021. This work was supported by the Cross Ministerial Strategic Innovation Promotion Program - automated driving for universal service Project entitled "Human Factors and HMI Research for Automated Driving." This article was recommended by Associate Editor Changxu Wu. *(Corresponding author: Huiping Zhou.)*

Huiping Zhou and Makoto Itoh are with the Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba 305-8573, Japan (e-mail: zhouhp@css.risk.tsukuba.ac.jp; itoh.makoto.ge@u.tsukuba.ac.jp).

Satoshi Kitazaki is with the Automotive Human Factors Research Center, National Institute of Advanced Industrial Science and Technology, Ibaraki 305-8566, Japan (e-mail: satoshi-kitazaki@aist.go.jp).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/THMS.2021.3051342.

Digital Object Identifier 10.1109/THMS.2021.3051342

time for an RtI that would allow a driver to respond effectively under different driving conditions, such as during noncritical and critical events and when performing nondriving-related tasks [18]–[20].

In such situations, one of the most important issues is determining the best way to maintain driver readiness to resume control from an automated driving system [21]. To improve the driver's perceptual ability regarding interventions under highly automated driving conditions, one of the more important issues is designing a human-machine interface (HMI) capable of providing a continuous feed of dynamic information to ensure that the user maintains situational awareness of evolving threats while the system is operating. These include visual cues and/or auditory messages [22], along with haptic indicators [23].

Although good experience and training are definitely efficacious for improving driver interventions [25], as automated cars become more widespread, it is inevitable that vehicles with automated driving systems will be controlled by drivers who have no experience with such systems. This makes it imperative to provide appropriate training. In their study, Payre *et al.* [26] focused on investigating the impact of training (mainly simple and elaborate *practice* in a simulator, as well as text and tutorial video instruction on automated driving) on driver interventions. In addition to empirically based learning (experience and training), explanation-based learning is another possible approach [27]–[29]. In explanation-based learning, which can be considered a general term for compiling knowledge [29], chunking allows new concepts to be deduced from existing concepts [30]. This indicates that good explanation-based knowledge has the potential to contribute to good information processing in the dynamic decision-making processes in highly automated driving systems [32], [33].

We focus on the study of human-machine interactions in the dynamic information processing of driver intervention to illustrate the concepts of explanation-based knowledge about SL. This article aims to demonstrate the necessity of such knowledge, to reveal why it is necessary, and to present how that knowledge influences driver take-over performance during conditional driving automation by collecting data on novice drivers with no experience in using any of the assisted/automated driving systems. Hence, the following research questions are proposed.

1) Is knowledge-based learning necessary?
2) What kind of knowledge is more serviceable to help drivers to perceive SL?
3) Could drivers extrapolate explained typical scene(s) to unexplained scenes?
4) Could the learning effect compensate for the differences due to the different extents of instructed knowledge?

II. Framework of HMI in the Information Process of Taking Over Control

Regarding decision making and knowledge, the most important part of the mental model is thought to be a cognitive structure consisting of specific knowledge and experience corresponding to a specific behavior modification [33]–[35]. This could represent knowledge and situation information that enhance human operators' understanding, reasoning, and prediction [36]–[38]. Hence, this study takes three concepts—decision making, mental model, and knowledge—into consideration within a framework illustrating information flow among the driving environment, system, human operator, and knowledge in the dynamic decision-making process of taking over control from conditional driving automation, as shown in Fig. 1.

In the learning theory of behavior modification, internal motivation, evocative stimuli, and situational stimuli are critical parts of the mental model [38]–[40]. *Internal motivation* is defined as goal-oriented arousal and a psychological force that leads an individual to process information, which is one of the crucial factors influencing drivers' behavior [41], [42]. The motivation in this study denotes an internal force that could stimulate drivers to resume vehicle control in the case system failures and/or surpassed limitations. *Evocative stimulus* is defined as an external incentive trigger to elicit a particular behavior, i.e., in our study, taking control of the vehicle. Hence, the RtI is thought to be an evocative stimulus. *A situational stimulus* denotes environmental information in a particular situation in which the SL is triggered.

When a driver employs conditional automation, they are not required to be aware of the dynamic situation when SL does not occur. However, internal motivation impacts the response-eliciting potency of behavior modification for a particular situation. Once SL occurs, an external stimulus of the RtI evokes an intentional response to the particular situation. By perceiving the situation with intention, the driver makes a reaction selection.

The explicit representation of knowledge is thought to be attributed to the mental model of motivational response in various fields [38], [42]–[44]. Hence, knowledge-based learning should correspond to *internal motivation*, *evocative stimuli*, and *situational stimuli*. Specifically, in a dynamic decision-making process of taking over car control due to SL, the possibility for the occurrence of SL provides a motivational intention for taking-over behavior. RtI provides the evocative message via the HMI and scenes related to the SL supply situational cues when the SL happens.

Since this study aims to discuss the extent to which knowledge affects drivers' taking over control, explanation-based knowledge is broken down into five levels in terms of the three concepts of occurrence of SL, RtI, and scene(s) related to SL. The instructed contents at each level are represented in Table I. Because novice drivers do not have any previous knowledge, experience, or training, it is supposed that their initial knowledge is determined by the instructed knowledge.

III. Method

A. Apparatus

A D^3sim fixed-base research and development driving simulation system (Mitsubishi Precision Co., Ltd., Tokyo, Japan) was used during the data collection phase of this study (see Fig. 2). This system provides simulated longitudinal direction control

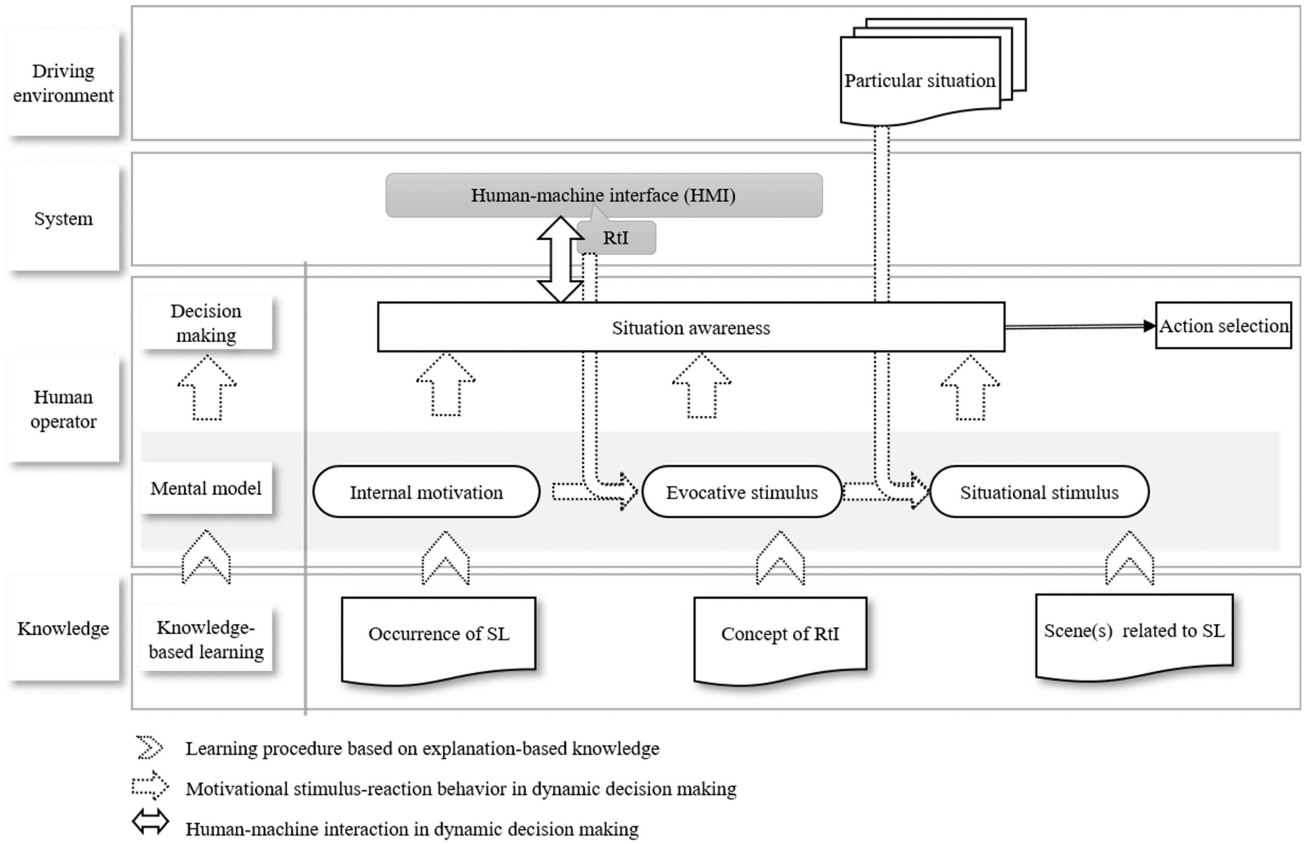


Fig. 1. Framework that illustrates information flow among the driving environment, system, human operator, and knowledge in the dynamic decision-making process for intervening in automated driving control due to system limitations.

TABLE I
FIVE LEVELS OF EXPLANATION-BASED KNOWLEDGE (LOK) ACCORDING TO HMI IN DYNAMIC DECISION MAKING

LoK	Contents in knowledge-based learning			
	Occurrence of SL	Concept of HMI	Scene(s) related to SL	Instruction about SL
0	No	No	No	None
I	Yes	No	No	Driver should only take over the vehicle's control once system failure/limitation occurs.
II	Yes	Yes	No	Driver should take over the vehicle's control once system limitation occurs, at which time an RtI will be given to the driver via the HMI.
III	Yes	Yes	Yes (Partial)	Driver should intervene in the vehicle's control once system failure/limitation occurs. An RtI will be issued to the driver via the HMI when the failure/limitation occurs and the system is unable to maintain control safely (for example, lane closed due to falling objects and bad weather).
IV	Yes	Yes	Yes (All)	Driver should intervene in the vehicle's control once system failure/limitation occurs. An RtI will be issued to the driver via the HMI when the failure/limitation occurs and the system is unable to maintain control safely in the following scenes: lane closed due to falling objects, bad weather, expressway's junction, blurred lane marker, vehicle intrusion from the right adjacent lane.

*Note that pictures of real-life scenarios were used for explaining these situations in which system limitation occurs.

Fig. 2. Fixed-base driving simulator operated by a R&D driving simulation system and electric control loading systems.

(acceleration and brake pedals) and electric control loading systems (MOOG Inc., Elma, NY, USA) to simulate lateral direction (steering wheel) control on a multilane expressway.

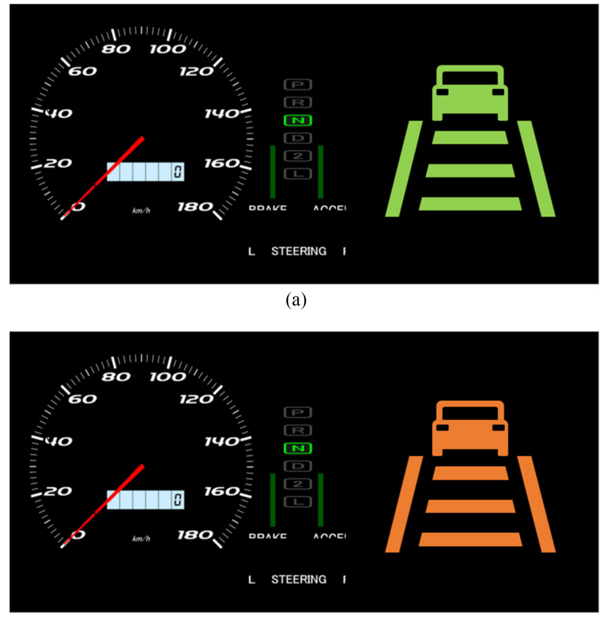
The vehicle control installed in the driving simulator was designed to operate at SAE Level 3. More concretely, the automated driving system manipulates the longitudinal (e.g., target following, cruising) and lateral (e.g., lane keeping) directions during vehicle control. Note that automatic lane changing was not operative in the conditional driving automation used in this study.

Conditional driving automation allows drivers to release their hands from the steering wheel and remove their feet from the accelerator and brake pedals when automated driving is engaged. However, if an RtI is issued, automated driving will be cancelled 10 s later, and the driver must resume full control of the vehicle by that time. Merat *et al.* [18] and Melcher *et al.* [19] proved that a lag of 10 s was sufficient for the driver to resume vehicle control comfortably and safely.

Fig. 3(a) shows the automated driving system's human–machine interface as it appears to the drivers. When an RtI is issued, the HMI screen picture changes to orange and blinks at 5 Hz simultaneously with an audible "beeping" signal, as shown in Fig. 3(b).

B. Participants

After receiving approval from the University of Tsukuba Research Ethics Committee (2016R119), 100 test subjects ranging in age from 19 to 85 years old (mean = 50.1, SD = 20.1) were recruited via a local organization as test participants. The distribution of the participants is described in Table II. Each participant possessed a valid driving license and drove daily. Note that special efforts (driver's self-assessment and past record of participating in experiments) were made to ensure that none of the test subjects had any prior knowledge of automated driving systems, including any experience using an autonomous car and any specific technical knowledge. After receiving an explanation of the data collection process, all test subjects gave written informed consent for their participation.



(a)

(b)

Fig. 3. (a) HMI when the automated driving system is active. (b) HMI of RtI including acoustic alert message and blinking visual icon has been issued.

TABLE II
NUMBER OF PARTICIPANTS, AVERAGE AGE OF EACH GROUP

LoK	N		Age (Mean ± SD*)	
	Male	Female	Female	Male
0	8	12	47.1 ± 20.1	53.8 ± 25.3
I	8	12	48.5 ± 12.9	58.3 ± 20.5
II	8	12	44.6 ± 17.7	58.4 ± 21.9
III	9	11	46.2 ± 19.9	59.6 ± 18.5
IV	9	11	50.9 ± 17.7	55.9 ± 22.8

* SD = Standard deviation.

C. Tasks

After starting the vehicle, drivers were asked to perform two tasks in relation to the automated driving system.

1) Driving Safely: Each driver was instructed to operate the simulated vehicle as *safely* as they would a real vehicle on an expressway prior to engaging the automated driving system, after which they could release their feet from the pedals and remove their hands from the steering wheel.

2) Nondriving-Related Task: Note that since nondriving-related tasks can be performed in cases involving conditional driving automation, participants were instructed to use the standardized visual surrogate reference task (SuRT) [45], [46]. Specifically, each participant was instructed to begin the SuRT upon hearing an audible signal that was given after the start of the automated driving system.

D. Explanation-Based Knowledge

In our study, the explanation-based knowledge consisted of 1) the automated driving system user's manual, i.e., information on the lateral and longitudinal control and system operation procedures; and 2) system failure/limitation. There are five levels of explanation-based knowledge about system failure/limitation, as represented in Table I.

E. Scenes

According to SAE J3016 [4], functional system limitations are thought to involve geographic, roadway, environment, traffic, speed, and/or temporal functional errors. Thus, eight scenes were designed for the following eight scenarios in which SL occurs.

1) Car approaches expressway junction [#01_Junction].
2) Lane is closed because of falling objects (#02_Lane(a)], stopped vehicle(s) [#05_Lane(b)], or a construction site [#09_Lane(c)].
3) Range of visibility is below 40 m due to heavy fog (#04_Fog).
4) Lane markers are blurred (#08_Lanemark).
5) A neighboring car suddenly intrudes from an adjacent lane and closely approaches the host car (#10_Intrusion).
6) System function failure occurs (#11_Failure).

In addition, three scenes in which no system failure/limitation occurs were also prepared for three other trials (#03, #06, and #07). Note that the numbers in parentheses refer to the executed order of each trial.

In each trial, approximately 30 s after the start of automated driving, participants were instructed to start a nondriving-related task after hearing an audible signal. All trials lasted an average of 174 s (SD = 38 s), and events occurred approximately 160 s after starting the drive. This signal was different from the acoustic message of RtI. Note that the meanings of the two types of audible messages were clearly explained to participants.

F. Experimental Design

Our experiment employed a single-factor design. The single factor is the level of knowledge of the SL. A single-factor (with five levels) analysis of variables (ANOVA) was performed for all variables. The factor was between subjective. A significance level of $p = 0.05$ was used.

G. Procedure

It must be noted that this study strictly controls the instructions of the SL. First, to reduce individual reading variations, an experimenter read the instruction manual of the driving automation system page-by-page while showing it through a display to each participant. It should also be noted that the experimenter was required to obtain the participant's approval before proceeding to the next page. However, to eliminate potential effects of variations in the experimenter's commentary on the limitations of the automated driving system, the experimenter was not permitted to answer any questions or provide additional clarifications to the reading material.

All participants were divided randomly into five groups according to the five SL explanation levels: *none*, *feasibility, HMI, partial,* and *all-scenes*. Participants received an explanation of the experiment's primary purpose and driving tasks. They signed an informed consent form after agreeing to participate in the experiment. After being introduced to the driving simulator, participants were given approximately five minutes of manual driving time to familiarize themselves with the device. After the automated driving system's user manual was read to all participants, two three-minute driving trial exercises were conducted in which every driver was required to intervene in vehicle control more than once. After confirming that a driver had successfully and smoothly taken over vehicle control from the system, the experimenter introduced the SuRT nondriving-related task.

Next, knowledge of system limitations was provided to each group at different levels.

Then, in the actual experimental step that followed, all drivers participated in eight event trials and three dummy trials, which were conducted in the same order from #01_Junction to #11_Failure. A maximum of one RtI was issued during each trial, and each participant was given a five-minute rest period after undergoing four trials.

Finally, each of the participants was interviewed.

H. Measures

To investigate how effective explanations for system limitations are in driver interventions, rates of successful intervention (RSI) and driver response times (RT) in response to RtIs were collected. This study defines successful interventions as cases in which the driver resumes full control of the vehicle safely within the 10 s period following the RtI and, thus, before automated driving control is completely cutoff. Because RTs are used to interpret the driver's perception, recognition, and decision to physically respond [47], [48], this study denotes it as the time from the moment that the RtI is issued to the moment that the driver initiates the intervention—either by operating the steering wheel, i.e., the operated steering angle is larger than 30° or depressing the accelerator/brake pedal, i.e., the slot opening is larger than 0.3. These variables are calculated based on log data recorded by the driving simulator.

The other two dependent variables used were the standard deviation of the lateral position (SDLP) and maximum steering angular velocity, which were calculated for 15 s from the RtI at 30 Hz.

IV. Results

A. Rate of Successful Intervention

The results of ANOVA conducted on the RSI for each participant show that the effect of the SL explanation level is statistically significant (F (4, 95) = 7.69, $p < 0.001^{**}$). According to Tukey's test, the significant differences between any pair of *LoK-0* versus *III* ($p < 0.001^{**}$), *LoK-0* versus *IV* ($p < 0.008^{**}$), *LoK-I* versus *III* ($p < 0.001^{**}$), and *LoK-I* versus *IV* ($p < 0.019^{**}$) [see Fig. 4(a)] indicate that the driver's interventions under *LoK-III* are more successful (*LoK-III*, RSI = 0.79 ± 0.04) when compared with the other four conditions (*LoK-0*, RSI = 0.49 ± 0.28; *LoK-I*, RSI = 0.51 ± 0.26; *LoK-II*, RSI = 0.67 ± 0.21, and *LoK-IV*, RSI = 0.72 ± 0.18). This result answers questions

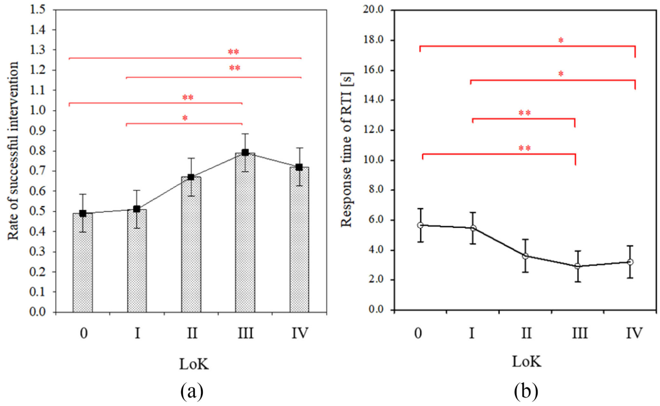


Fig. 4. (a) Mean rate of successful intervention (Error bars = standard error). (b) Response time of responding to request to intervene (RT) (Error bars = standard error).

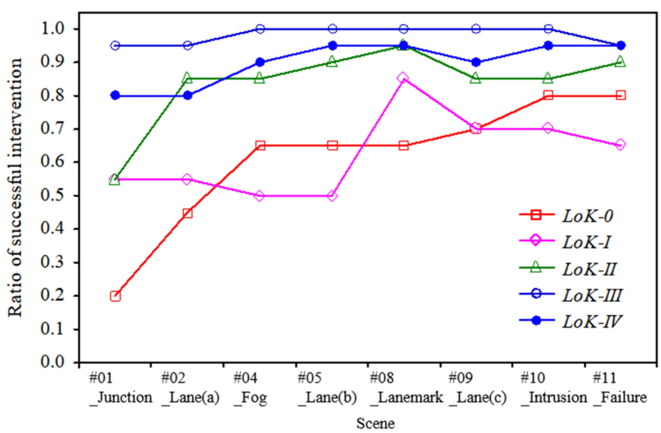


Fig. 5. Mean rate of successful intervention under five levels of explanation on system limitation as a function of the experienced trial.

(i) and (ii); i.e., knowledge is necessary, and sample scenes are more serviceable to help drivers comprehend SLs.

Fig. 5 shows RSIs from responding to RtIs as a function of experienced trials. In addition to *LoK-I*, the rate increases progressively as experience is gained. This overwhelming increase from #01_Junction to #02_Lane (a), particularly under conditions of *LoK-0* and *LoK-II*, indicates a learning effect of experience different from that of instructed knowledge.

The results also show that the rate at *LoK-III* maintains a value above 90%, which is even better than *LoK-IV*, and that even though the success rate at *LoK-0* grows most significantly, the highest rate achieved does not exceed that shown in Fig. 4.

B. RT to the RtI

Fig. 4(b) shows the average value and standard errors of the RT response to RtI as a function of SL explanations. According to the ANOVA conducted on the RT, the primary effect of the SL explanations is significant (F (4, 91) = 5.79, $p < 0.001^{**}$). Tukey's test revealed significant differences between any pair of *LoK-0* versus *LoK-III* ($p < 0.004^{**}$), *LoK-0* versus *LoK-IV* ($p < 0.018^{*}$), *LoK-I* versus *LoK-III* ($p < 0.008^{*}$), and *LoK-I* versus *LoK-IV* ($p < 0.028^{*}$).

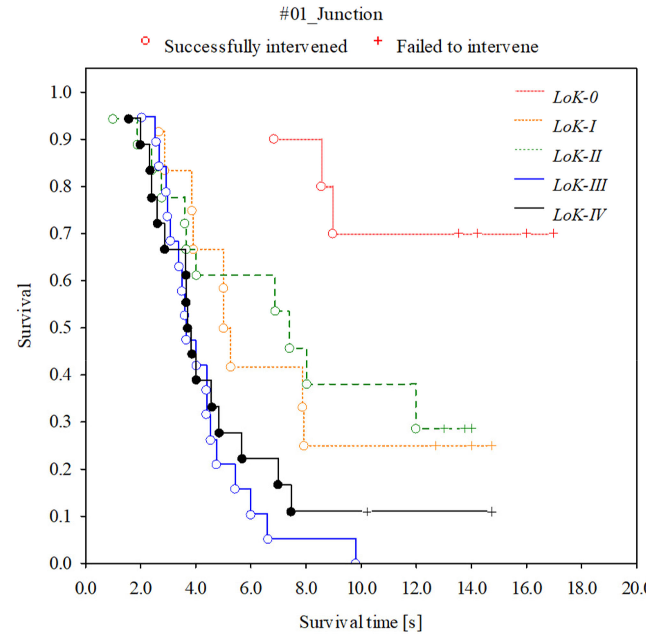


Fig. 6. Kaplan–Meier survival time in RSI under five levels of explanation on system limitation in #01_Junction.

C. Driver Behavior in the First Experienced Scene, #01_Junction

Kaplan–Meier survival curves for the five SL explanation levels in the full original data are shown in Fig. 6. According to the survival analysis, the effects of explanations of SL were significantly different at each of the five levels ($\chi^2 = 21.32$, $df = 4$, $p < 0.001^{**}$). Using the estimated survival curves, the absolute reduction in the probability of survival to 10 s from RtI was 70.0%, 25.0%, 38.9%, 0%, and 11.1% for *LoK-0*, *LoK-I*, *LoK-II*, *LoK-III*, and *LoK-IV*, respectively. Here, again, it should be noted that automated control is cancelled completely in cases where no intervention is initiated after 10 s have passed from the RtI.

D. Influence of Experience versus Knowledge

Drivers' interventions in similar scenes [#02_Lane(a), #05_Lane(b), and #09_Lane(c)] were investigated to determine the impact of experience-explanation knowledge on similar situations. In such cases, the important point was that the driver should change lanes after intervening in the vehicle's control to avoid a collision with a stopped vehicle. Note that it was only scene #09_Lane(c) in which the host vehicle was following a target, i.e., a driver could not visually recognize the construction site when the system issued an RtI.

ANOVA of the RT in response to RtI for explanation and experience (the three trials) is shown in Fig. 7(a). The effect is revealed in the RtI for both the SL explanations (F (4, 242) = 3.17, $p < 0.014^{*}$) and experience (F (1, 242) = 4.07, $p < 0.018^{*}$). According to Tukey's test, a significant difference could be observed between any pair of *LoK-0* versus *LoK-III* ($p < 0.031^{*}$). In addition, ANOVA was conducted for the time from the RtI to the lane change employed to examine lane-changing

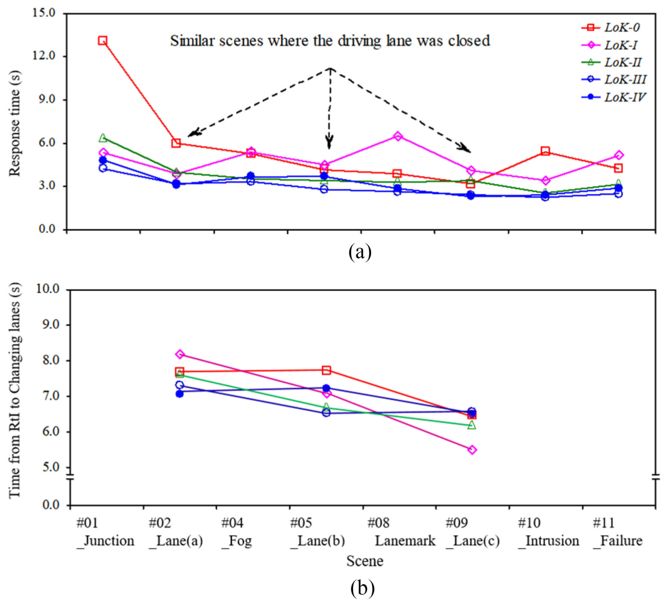


Fig. 7. Mean values of RT response to RtI, (a) and time from RtI to lane change (b) as a function of experienced trial where changing lanes was required in #02_Lane(a)), #05_Lane(b), and #09_Lane(c).

behavior after taking over. Note that the effect of experience is only significant (F (1, 231) = 13.14, p < *0.001***) for the time from the RtI to the lane change and had no effect on the SL explanations [F (4, 231) = 0.78, p = 0.534, see Fig. 7(b)]. The results statistically demonstrate the learning effect on driver interventions. Hence, we give an affirmative answer to Question (iii).

E. Driver Interventions in Ungiven Scenes

Note that *LoK-II, LoK-III,* and *LoK-IV* differ based on the drivers' knowledge about the related scenes. Among all the scenes used in this study, instructions for the scenes in #08_Lanemark (i.e., blurred highway lane pavement) and #10_Intrusion (i.e., sudden vehicle intrusion from adjacent lanes) were provided to participants in the *LoK-IV* group but not in the *LoK-II* and *LoK-III* groups. Fig. 8 shows the maximum steering wheel angular velocity after the driver's intervention.

After conducting a statistical comparison of these driving performance values, no significant difference was found (F (2, 57) = 0.11, p < 0.892) in #08_Lanemark, although a significant difference was reported for the maximum steering lateral angular velocity (F (2, 57) = 6.20, p < 0.007**) in #10_Intrusion. This result suggests that drivers would also deal well with unfamiliar scenes if they could effectively process lessons learned from other given scenes, which gives a positive answer to Question (iv).

F. Driver's Intervention During System Function Failure

A system function failure occurs in #11_Failure, in which it is difficult for drivers to perceive a hazard because the event is not obvious. According to ANOVA on RtI and SDLP, the primary effect of the SL explanations is revealed in driver intervention

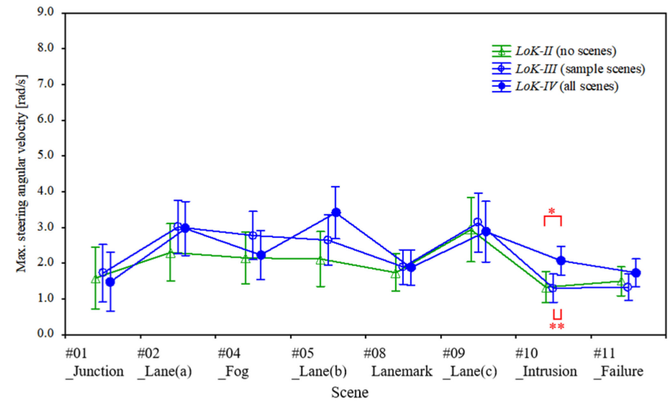


Fig. 8. Maximum steering angular velocity after intervention in all 8 trials, where #08_Lanemark and #10_Instrusion were not included in those sample scenes at *LoK-III*) (Error bars = standard error).

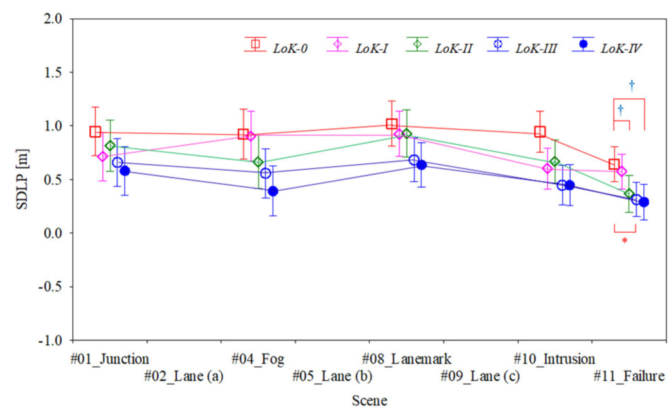


Fig. 9. SDLP during 15-s from RtI in the 5 trials (i.e., #01_Junction, #04_Fog, #08_Lanemark, #10_Instrusion and #11_Failure), where changing lanes was not required. (Error bars = standard error).

(F (4, 86) = 2.57, p <*0.05**). Tukey's test revealed a significant difference between *I* (possible of occurrence) and *III (partial scenes)* (p < *0.05**) (see Fig. 7).

Likewise, the ANOVA results also indicate the primary effect of the system limitation explanations on driver performance (SDLP, F (4, 86) = 2.57, p <*0.05**). Tukey's test showed that a significant difference (p < *0.05**) exists between the pair of *None* versus *(3)-Partial,* and that a slight difference exists between any pair of *LoK-0* versus *LoK-II* (p < *0.1*†) and *LoK-0* versus *LoK-IV* (p < *0.1*†) (see Fig. 9).

G. Limitations

Because this study was conducted in a fixed-base driving simulator, drivers might pay more attention to the nondriving-related task than they would when driving a real vehicle. This might lead to a relatively low RSI, especially in the first scenario (#01_Junction).

Another limitation of this study is that an order effect was not removed. We intentionally fixed the scenes' order to focus on the effects of knowledge and its variation with the experience at different periods, as shown in Fig. 5. Consequently, we must admit that the results might be affected by order, as shown in

Fig. 4. On the other hand, although we controlled the scenes' effect to focus only on discussing experience influence, it is difficult to ignore the effect of the scene type.

Finally, this study did not analyze drivers' gaze behavior before and after taking over car control, which is an important way to measure drivers' attitudes toward an automated system. However, we could not conduct statistical analysis on gaze behavior due to a failure in collecting eye movement.

V. Discussions and Implications

Overall, our experimental results demonstrate the necessity of knowledge-based learning for novice drivers and experienced drivers without appropriate knowledge, such as people in *LoK-0* and *I*.

Although the drivers in *LoK-II* (who were instructed in the concept of the RtI and the possible occurrence of limitations) intervened in vehicle control from automated driving more safely than those in *LoKs 0* and *I*, their performance was still worse than those in *LoKs III* and *IV*. The results revealed the importance of instructing the concept of RtI in HMI.

Pazzani's [30] study implied that causal knowledge could dominate the influence of an individual's logical form, which might explain why *LoKs III* and *LoK-IV* performed better than the others. More specifically, the knowledge under conditions of *LoKs III* and *IV* presents the related scenes that would support the driver in perceiving a situational stimulus for comprehending the causal relationship between the issued RtI and the given scene. This causal representation helps drivers process information more logically and, thus, perceive potential hazards more thoroughly.

On the other hand, from the ecological perspective of human information processing, a driver's response to the incoming stimulus is addressed as perception-action coupling [49], [50] in which RT reflects strategic aspects of coding rules or mental representation [51], [52]. The typical scenes taught through explanation-based knowledge represented the decision rules of the SLs. The largely decreased RT under *LoK-III* (partial scenes) demonstrates that strategically acquiring knowledge about these rules significantly improves the functionality of the mental model when reacting to an issued RtI (i.e., evocative stimulus) in a particular situation (i.e., situational stimulus). That is, the mental model is serviceable to the driver's stimulus action when responding to a critical situation.

It should also be noted that the participants in the *LoK-III* group were only instructed on some of those scenes. Nevertheless, these drivers' taking-over performance was achieved in the same manner as those in the *LoK-IV* group. This suggests that drivers might be able to internally generate specific scene(s) in response to general scenes. Moreover, if related scenes are presented in an adequate way, drivers are still able to achieve successful intervention in an ungiven critical scene (such as *#08_Lanemark*, *#10_Intrusion*, see Fig. 8) or a perplexing urgent situation (such as *#11_Failure*, see Fig. 9).

This article also demonstrates the positive impact of experience on driver interventions (see Figs. 5 and 7). Nevertheless, a definite margin still exists between any combination of *LoK 0, I* or *II* and *LoKs III* and *IV*. Note that their differences imply that empirically based learning without explanation-based learning is limited in its capacity for improving driver intervention. The experimental results also demonstrate that a driver with an inadequate level of knowledge might still fail to accomplish an intervention even if they experience multiple related scenes. This finding indicates that practical experience could improve drivers' perceptual ability to observe external stimuli such as evocative stimuli through HMI and situational stimuli from the driving environment and that knowledge of SL occurrence could effectively induce internal motivation.

As mentioned in our study's limitations, driver reactions might be affected by scene type, such as the extent of danger, suddenness and perceivability of a hazard. Although the same type of SL occurred in #02_Lane (a), #05_Lane (b), and #09_Lane (c), it was difficult to identify what was happening while an RtI was issued in #09_Lane (c). Drivers' anxiety about an unpredictable situation might induce a more rapid reaction to the RtI, as shown in Fig. 7. Unidentified internal motivation relying only on experience also increased anxiety, resulting in a relatively rapid response, as revealed in Fig. 9(a), while a higher SDPL in Fig. 9(b) represents an unstable performance after taking over control from the system.

This study naturally presumes that a driver would be sufficiently motivated to take over the system's control if they could predict that the system would soon become inoperative due to failures or surpassed limitations in protection motivation theory [53]. However, in uncertainty processing theory, maintaining an adequate level of motivation requires some degree of uncertainty [54]. Giving partial scenes in *LoK-III* allows for the uncertainty that other scenes that are not explained beforehand might cause. This explains why the driver take-over performance in *LoK-III* appears to occur more stably and effectively, as shown in Fig. 4. In other words, the decreased uncertainty at a relatively low level in *LoK-IV* might have degraded the level of motivation.

When reviewing the comments from all participants, we found that drivers in *LoK-II* gave more negative comments (e.g., "the volume of the audio message is too low to hear") and fewer positive comments (e.g., "the message is useful to know what is happening") on the concept of HMI. This may be because those drivers depended strongly on information supplied through the HMI when a system failure or surpassed limitation occurred; therefore, they tended to expect more information from the HMI. Those drivers also showed a preference for a voice message over the simple audible signal used in our experiment. Such initiative comments about the HMI imply that more information might be expected from a *voice* message if no other knowledge source is available.

Finally, through the interview comments, drivers in *LoK-0* showed an overwhelming negative attitude of distrust toward automation (e.g., "I could not completely trust the system") but expressed relatively high expectations in relation to automated driving itself. The conflict between negative attitudes and positive expectations suggests that a lack of primary knowledge would influence driver trust calibration in the proper manner.

VI. Conclusion

This study investigated the influence of SL explanations on driver behaviors when responding to RtI under conditioned driving automation (SAE Level 3) conditions. Because drivers using conditional automation are not requested to monitor the operational environment, the instructed knowledge greatly impacts the drivers' take-over performance, as shown in this study. The experimental results presented in this article show that driver interventions could be degraded if SL explanations were insufficient in terms of the related system failures and/or surpassed limitations.

Additionally, we have shown that supplying some typical scenes in which system limitations might occur is essential for a driver's ability to generalize selection rules when working to perceive system failures and surpassed limitations, which supports the findings of Payre *et al.* [26]. This generalization ability is beneficial when it is necessary to respond to an unfamiliar situation.

Furthermore, the results of our participants' interviews appear to indicate that improving generalization abilities might also serve to facilitate driver trust and/or more positive attitudes in relation to automated driving systems.

Finally, even though this article clarifies the extent of knowledge regarding system failures and/or surpassed limitations that should be explained to a driver, the method used to supply that knowledge is still one of the more important issues related to driver intervention in a highly automated driving system; thus, it must be considered an important topic for our future work.

Acknowledgment

The authors would like to thank N. Kashiwazaki, Y. Noguchi, R. Taguchi, Y. Ishitsuka, S. Funoki, K. Toyoda, T. W. Liu, S. Nishina, and R. Umeno for their assistance in data collection and analysis.

References

[1] B. E. Sabey and H. Taylor, *The Known Risks We Run: The High Way*. New York, NY, USA: Springer, 1980, pp. 43–70.

[2] R. R. Knipling, "IVHS technologies applied to collision avoidance: Perspectives on six target crash types and countermeasures," in *Proc. Annu. Meeting IVHS Amer: Surf. Transp., Mobility, Technol., Soc.*, 1993, pp. 249–259.

[3] D. W. Eby and L. P. Kostyniuk, "Driver distraction & crashes: An assessment of crash data-bases & review of the literature," Univ. Michigan, Transp. Res. Inst., Ann Arbor, MI, USA, Tech. Rep. UMTRI-2003-12, 2003.

[4] S. A. E. International, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," *J3016, SAE Int.*, 2016.

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

[6] M. N. Lees and J. D. Lee, "The influence of distraction and driving context on driver response to imperfect collision warning systems," *Ergonomics*, vol. 50, no. 8, pp. 1264–1286, 2007.

[7] M. S. Young and N. A. Stanton, "What's skill got to do with it? Vehicle automation and driver mental workload," *Ergonomics*, vol. 50, no. 8, pp. 1324–1339, 2007.

[8] T. Inagaki and M. Itoh, "Human's overtrust in and overreliance on advanced driver assistance systems: A theoretical framework," *Int. J. Veh. Technol.*, vol. 2013, 2013, Art. no. 951762.

[9] D. Abbink, M. Mulder, and E. R. Boer, "Haptic shared control: Smoothly shifting control authority?," *Cogn., Technol. Work*, vol. 14, no. 1, pp. 19–28, 2012.

[10] R. Parasuraman and D. H. Manzey, "Complacency and bias in human use of automation: An attentional integration," *Hum. Factors*, vol. 52, pp. 381–410, 2010.

[11] M. Mulder, D. A. Abbink, and E. R. Boer, "Sharing control with haptics: Seamless driver support from manual to automatic control," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 54, no. 5, pp. 786–798, 2012.

[12] T. Radke, L. Lewandowitz, F. Gauterin, and T. Kaczmarek, "Operationalisation of human drivers' subjective dynamics perception in automated longitudinally controlled passenger vehicles," *Int. J. Veh. Auton. Syst.*, vol. 11, no. 1, pp. 86–109, 2013.

[13] N. Merat and A. H. Jamson, "How do drivers behave in a highly automated car," in *Proc. 5th Int. Driving Symp. Hum. Factors Driver Assessment, Training Veh. Des.*, 2009, pp. 514–521.

[14] A. H. Jamson, N. Merat, O. M. Carsten, and F. C. H. Lai, "Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions," *Transp. Res. C, Emerg. Technol.*, vol. 30, pp. 116–125, 2013.

[15] M. Hjālmdahl, S. Krupenia, and B. Thorslund, "Driver behavior and driver experience of partial and fully automated truck platooning-a simulator study," *Eur. Transport Res. Rev.*, vol. 9, no. 8, Aug. 2017.

[16] K. Zeeb, A. Buchner, and M. Schrauf, "What determines the take-over time? An integrated model approach of driver take-over after automated driving," *Accident Anal. Prevention*, vol. 78, pp. 212–221, 2015..

[17] T. Mioch, L. Kroon, and M. A. Neerincs, "Driver readiness model for regulating the transfer from automation to human control," in *Proc. 22nd Int. Conf. Intell. User Interfaces*, 2017, pp. 205–213.

[18] N. Stanton, M. Young, I. Karlsson, and L. Nilsson. "Semi-automated versus highly automated driving in critical situations caused by automation failures," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 27, pp. 218–228, 2014.

[19] N. Merat, A. H. Jamson, F. C. H. Lai, M. Daly, and O. M. J. Carsten, "Transition to manual: Driver behaviour when resuming control from a highly automated vehicle," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 27, pp. 274–282, 2014.

[20] V. Melcher, S. Rauh, F. Diederichs, H. Widlroither, and W. Bauer, "Take-over requests for automated driving," *Procedia Manuf.*, vol. 3, pp. 2867–2873, 2015.

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

[22] N. F. and A. Nekum, "Timing of in-vehicle advisory warnings based on cooperative perception," in *Proc. Hum. Factors Ergonom. Soc. Eur. Chapter Annu. Meeting*, 2014, pp. 1–13.

[23] S. M. Petermeijer, J. C. de Winter, and K. J. Bengler, "Vibrotactile displays: A survey with a view on highly automated driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 897–907, Apr. 2016.

[24] J. D. Rupp and A. G. King, *Autonomous Driving - A Practical Roadmap*. Warrendale, PA, USA: SAE Int., 2010.

[25] G. Maycock and C. R. Lockwood, "The accident liability of british car drivers," *Transport Rev.*, vol. 13, no. 3, pp. 231–245, 1993.

[26] W. Payre, J. Cestac, N. Dand, F. Vienne, and P. Delhomme, "Impact of training and in-vehicle task performance on manual control recovery in an automated car," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 46, pp. 216–227, 2017.

[27] R. Michalski, "A theory and methodology of inductive learning," in *Machine Learning: An Artificial Intelligence Approach*, R. Michalski, J. Carbonell, and T. Mitchell, Eds. San Mateo, CA, USA: Morgan Kaufmann, 1983, pp. 83–134.

[28] T. Mitchell, S. Kedar-Cabelli, and R. Keller, "Explanation-based learning: A unifying view," *Mach. Learn.*, vol. 1, pp. 47–80, 1986.

[29] M. J. Pazzani, "Influence of prior knowledge on concept acquisition: Experimental and computational results," *J. Exp. Psychol., Learn., Memory, Cognitions*, vol. 17, no. 3, pp. 416–432, 1991.

[30] J. R. Andenon, "A theory of the origins of human knowledge," *Artif. Intell.*, vol. 40, pp. 313–351, 1989.

[31] J. E. Laird, N. A., and P. S. Rosenbloorn, "SOAR: An architecture for general intelligent," *Artif. Intell.*, vol. 33, pp. 1–64, 1987.

[32] E. R. Boer and A. Liu, *Cambridge Basic Research 1997 Annual Report, Nissan CBR*. Cambridge, MA, USA, 1997, Art. no. 02142.

[33] E. R. Boer, E. C. Hildreth, and M. A. Goodrich, "A driver model of attention management and task scheduling: Satisficing decision making with dynamic mental models," in *Proc. 17th Eur. Annu. Conf. Human Decis. Making Manual Control*, Valenciennes, France, Dec. 14–16, 1998.

[34] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, “A model for types and levels of human interaction with automation,” *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 30, no. 3, pp. 286–297, May 2000.

[35] A. J. Elliot and M. V. Covington, “Approach and avoidance motivation,” *Educ. Psychol. Rev.*, vol. 13, no. 2, pp. 73–92, Jun. 2001.

[36] P. N. Johnson-Laird, “Mental models,” in *Foundations of Cognitive Science*. Cambridge, MA, USA: MIT Press, 1989.

[37] W. B. Rouse and N. M. Morris, “On looking into the black box: Prospects and limits in the search for mental models,” *Psychol. Bull.*, vol. 100, no. 3, pp. 349–363, Nov. 1986.

[38] D. Gentner, “Mental models, psychology of,” in *International Encyclopedia of the Social and Behavioural Sciences*. H. J. Smelser and P. B. Bates, Eds. New York, NY, USA: Elsevier, 2002, pp. 9683–9687.

[39] D. Bindra, “A motivational view of learning, performance, and behavior modification,” *Psychol. Rev.*, vol. 81, no. 3, pp. 199–213, 1974.

[40] M. Blaskova, M. Bizik, and R. Jankal, “Model of decision making in motivating employees and managers,” *Inzinerine Ekonomika-Eng. Econ.*, vol. 26, no. 5, pp. 517–529, 2015.

[41] J. Kinard, *Management*. Lexington, MA, USA: D.C. Heath Company, 1988.

[42] J. R. Bettman and C. W. Park, “Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis,” *J. Consum. Res.*, vol. 7, no. 3, pp. 234–248, 1980.

[43] A. B. Markman, *Knowledge Representation*. New York, NY, USA: Taylor & Francis, 1999.

[44] G. R. Jones and J. M. George, *Essentials of Contemporary Management*. New York, NY, USA: McGraw-Hill, 2014.

[45] ISO14198, “Road vehicles - Ergonomic aspects of transport information and control systems - Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems,” 2012, ISO/TS 14198:2012(en).

[46] M. S. and H. A., “Surrogate distraction measurement techniques: The lane change test,” in *Diver Distraction Theory, Effects, and Mitigation*, M. R, J. Lee, K. Young, Eds. Boca Raton, FL USA: CRC Press, 2008, pp. 107–122.

[47] P. L. Olson, “Driver perception response time,” *SAE Tech. Papers*, 1989, Art. no. 890731.

[48] T. Magister, M. Batista, and L. Bogdanović, “Measurement of the driver response time in the simulated and real emergency driving situations,” *PROMET-Traffic Transp.*, vol. 18, no. 1, pp. 23–32, 2006.

[49] T. Hasbroucq, Y. Guiard, and L. Ottomani, “Principles of response determination: The list-rule model of SR compatibility,” *Bull. Psychonomic Soc.*, vol. 28, pp. 327–330, 1990.

[50] P. R.W., T. Van Zandt, C. Lu, and D. J. Weeks, “Stimulus-response compatibility for moving stimuli: Perception of affordances or directional coding?,” *J. Exp. Psychol., Human Perception Perform.*, vol. 19, pp. 81–91, 1993.

[51] C. F. Michaels, “Destination compatibility, affordances, and coding rules: A reply to Proctor, Van Zandt, Lu, and Weeks,” *J. Exp. Psychol., Human Perception Perform.*, vol. 19, 1993, Art. no. 1121-l127.

[52] J. Alegria and P. Bertelson, “Time uncertainty, number of alternatives and particular signal-response pair as determinants of choice reaction time,” *Acta Psychologica,* vol. 33, pp. 33–44, 1970.

[53] R. W. Rogers, “Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation,” in *Social Psychophysiology*, J. Cacioppo, R. Petty, Eds. New York, NY, USA: Guilford Press, 1983.

[54] P. Anselme, “The uncertainty processing theory of motivation,” *Behav. Brain Res.*, vol. 208, no. 2, pp. 291–310, 2010.