Int J Interact Des Manuf
DOI 10.1007/s12008-014-0227-2

ORIGINAL PAPER

Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance

Jeamin Koo · Jungsuk Kwac · Wendy Ju · Martin Steinert · Larry Leifer · Clifford Nass

Received: 12 February 2014 / Accepted: 24 March 2014

Abstract This study explores, in the context of semi-autonomous driving, how the content of the verbalized message accompanying the car's autonomous action affects the driver's attitude and safety performance. Using a driving simulator with an auto-braking function, we tested different messages that provided advance explanation of the car's imminent autonomous action. Messages providing only "how" information describing actions (e.g., "The car is braking") led to poor driving performance, whereas "why" information describing reasoning for actions (e.g., "Obstacle ahead") was preferred by drivers and led to better driving performance. Providing both "how and why" resulted in the safest driving performance but increased negative feelings in drivers. These results suggest that, to increase overall safety, car makers need to attend not only to the design of autonomous actions but also to the right way to explain these actions to the drivers.

Keywords Semi-autonomous driving · Feedforward alerts · Car–driver interaction

J. Koo (✉) · W. Ju · L. Leifer
Department of Mechanical Engineering, Center for Design Research, Stanford University, Stanford, CA, USA
e-mail: finepine@stanford.edu

J. Kwac
Department of Electric Engineering, Stanford University, Stanford, CA, USA
e-mail: kwjusu1@stanford.edu

M. Steinert
Department of Engineering Design and Materials, Norwegian University of Science and Technology, Trondheim, Norway

C. Nass
Department of Communication, Stanford University, Stanford, CA, USA

1 Introduction

Cars are starting to make decisions on our behalf. Functionalities that take over control from drivers such as adaptive cruise control, lane keeping, and self-parking systems are readily available from almost every auto manufacturer. Google is currently prototyping the operation of fully autonomous driving vehicles on public roads in the USA [1], and US states such as California, Nevada, Florida, and Michigan have started to approve a driving license for autonomous cars under new laws and regulations [2,3].

As autonomous driving capabilities become increasingly widespread, state-of-the-art sensing, vision, and control technologies will enable cars to detect and monitor every object around the car, relying on real-time object measurements. Additionally, in-vehicle information technology will be fully capable of delivering both external (terrain) and internal (machine) information about the car to a driver inside the cabin.

Connected cars, which are networked to traffic sensory and online information about road conditions, will transform the ways we perform in the driver's seat.

Given these dramatic changes in the user experience, designers are challenged to model appropriate ways of conveying necessary and timely information to the driver. Typical human–machine interactions focus on alerting the driver to the actions of the car with an in-vehicle warning system. Designers should review and reassess the interaction between car and driver.

Over the last hundred years of automobile history, primary control for driving has always belonged to a human driver. Despite a great many innovations in the *powertrain* domain—gasoline and diesel fuels, automatic transmission, electric and hybrid propulsion, etc.—human hands and feet have always remained in contact with the steering wheel and the pedals, the means of controlling the car. Cars that drive themselves, however, allow drivers to forgo physical manipulation of the steering wheel and pedals. This is a big departure from the driving paradigm of the last century. In this new era of control delegation, there is a compelling need for designers to observe this phenomenon from the perspective of the human driver. Hence, our research adopts a user-centered method towards designing autonomous car interactions; it is important that the design of autonomous car behaviors should be informed by and tested against studies of human behavior in the driving context. As Norman states [4], "We must design our technologies for the way people actually behave, not the way we would like them to behave." By running controlled studies of driver response to vehicle interactions, we are better able to predict user behavior.

1.1 Importance of automation representation

The challenge of designing automation is to better understand how the automation interacts with the human operator, and how the system's operational information is best delivered to the operator. Feedback plays an important role in car–driver interactions. As many researchers observe, systems typically lack the essential concept of "feedback" to let operators know what actions are occurring. Norman points out that the central problems in conveying the information are inadequate interaction and inappropriate feedback from the car (machine) to human (operator) [5]. Stanton and Young emphasize that feedback from the automated system is required to keep the driver up to date, and that feedback is one of the essential features when designing a semi-autonomous driver-supported system [6–8]. But feedback alone is not sufficient: without proper context, abstraction, and integration, feedback may not be understandable to the driver [9]. Enhanced feedback and representation can help prevent the problems associated with inadequate feedback, which range from driver mistrust to lack of driver awareness and difficulty in recovering from errors.

Our approach takes a different angle on providing information to drivers. The nature of feedback is to inform users of the direct outcome of the system's action. However, in autonomous driving scenarios, we claim that it is crucial to provide information to drivers ahead of the event (Fig. 1). Such "feedforward" information allows the driver to respond appropriately to the situation and to gain trust that the car is taking control for a good reason.

1.2 Understanding the user scenario

As human drivers relinquish their control over the car, it becomes increasingly important to understand how drivers, as users, perceive and accept the intelligent automated function. Imagine, for example, when a car senses obstacles and is about to brake automatically. On the one hand, the system could provide an alert regarding how the car is going to act. On the other hand, the system could supply information about why the car is going to perform that action. For instance, the warning could be "Car is braking" or "Obstacle ahead." The first message conveys the operational behavior that the car is about to initiate, whereas the second message conveys the driving context (environment) that the car is about to encounter. From a user perspective, the driver may be interested in both types of information: *how* the car will behave (longitudinally or laterally) and *why* the car is behaving that way.

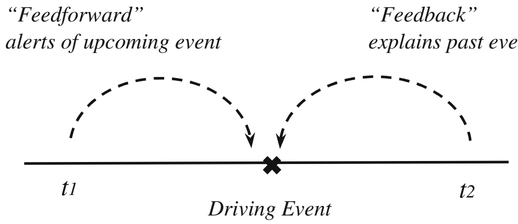

Fig. 1 The timing of the alert occurrence in a driving situation

Therefore, a key design question arises here: When it comes to informing drivers about impending autonomous behavior, how should we generate appropriate messages explaining the machine's intelligence and intention? The overarching goal of this paper is to explore user response to cars that communicate their automated action to the driver in different ways. In a simulation of diverse driving conditions, we deliver differently designed messages and then assess the consequences of the message design on driver attitude and performance.

Essentially, our design method is to develop vehicle interface and interaction designs that embody competing hypotheses for what will positively or negatively affect user perception and behavior, and to test these prototype designs in a simulator or controlled driving setting.

2 Method

2.1 Experiment overview

Our study explored two types of feedforward information: *how* the car is acting (what automated activity it is undertaking) and *why* the car is acting that way, as well as a combination of *how* and *why* messages. We employed a two-by-two between-participants experimental design (Table 1).

Within this study structure, we determined whether drivers benefit from the car explaining its actions instead of merely acting without explanation. In the experiment, the autonomous action consisted of the car automatically braking to prevent impending collision. In these situations, a voice

Table 1 Structure of the study: 2 (cars telling *why*: yes/no) × 2 (cars telling *how*: yes/no)

	Why message	Without *why* message
How message	Both internal and external information referring to automation	Internal (car activity) information
Without *how* message	External (situation) information	No information (control condition)

alert informed the driver of the car's imminent autonomous behavior. We tested three different message designs:

1. *How* message: Information about how the car is acting, announcing the automated action the car is initiating. In our experiment, this message was, "Car is braking."
2. *Why* message: Situational information explaining the reason for engaging automation: "Obstacle ahead."
3. *How* + *why* message: Alert of how the car is acting and why the car is making those actions: "Car is braking due to obstacle ahead."

We assessed driver attitudes and safety performance to learn how different types of information affect the driver.

The study consisted of two driving simulator sessions (training and data runs) for each participant, lasting approximately half an hour in total. After completing each driving session, participants answered an online questionnaire. The entire study was conducted in one room at Stanford University where both the driving simulator and the computers with the online questionnaire were located.

2.2 Participants

Sixty-four university students (32 males and 32 females, gender-balanced across each condition) with valid driver's licenses were recruited to participate in the study for course credit. They were aged 18–27 ($M = 21.11,\ SD = 1.42$) and had between two and ten years ($M = 4.99,\ SD = 1.55$) of driving experience. All participants gave informed consent and were debriefed after the experiment.

2.3 Apparatus

We used a driving simulation called STISIM from Systems Technology, Inc., which has been used for previous driving studies conducted by the Communication Between Humans and Interactive Media (CHIMe) Lab at Stanford University [10,11]. Physically, the simulator consists of a half-cut modified Ford Mustang equipped with a gas pedal and brake, a force-feedback steering wheel, and a driver's seat. During the experiment, the simulated driving course was run on lab computers and projected onto three rear-projection screens (each 2.5 m diagonal) angled so that the driver had a 160° field of view (Fig. 2). The same simulator setup was used for both the training course and the main driving course.

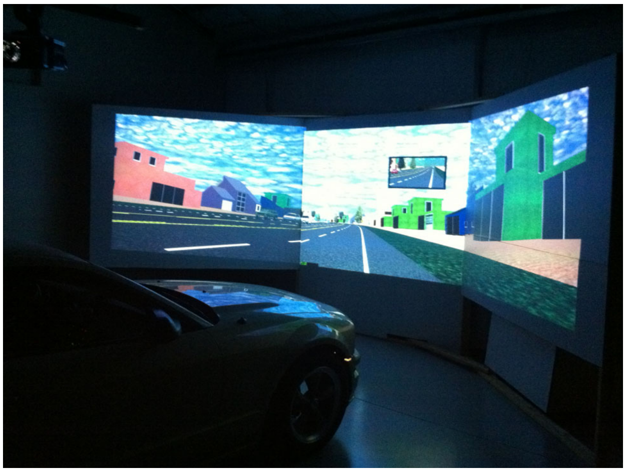

Fig. 2 Overview of the driving simulation setup

For the purpose of this experiment, we built and programmed an auto-braking function in the simulator and designed it to brake automatically in impending collision situations. Thus, the car was able to take control from the driver and decelerate when responding to unexpected circumstances that required emergency braking, such as an obstacle on the road or a pedestrian jaywalking. The participants were informed that the automated function would activate only for the purpose of safety support. To minimize effects caused by the lack of motion feedback that would exist in a real driving situation, a braking sound was simultaneously provided as a cue of decelerating action every time the braking force was applied. Additionally, to better simulate the driving experience, force feedback on the brake pedal was generated in proportion to the brake pressure. These two features acted as physical proxy for braking action.

The 12-km driving course incorporated urban, suburban, and highway sections, and featured stop signs and traffic signals. The course included several hazards, traffic variations, environmental scenery, and changing driving conditions to mimic an actual difficult driving experience. Speed limit varied over the course from 30 miles per hour in urban areas to 65 miles per hour on the highway. The course was specifically designed to avoid motion sickness in participants.

In the experiment, whenever an unexpected challenge appeared on the course, the voice warning and/or auto braking was generated. For example, as soon as debris or a jaywalker suddenly appeared in the middle of the road, the car responded to the unexpected event by generating a voice alert and applying the brake for the driver. For the voice alert, a standard American accent with no particular mood or inflection was employed, and the alert's latency time of response to the automated braking was approximately one second. Participants were informed that the alerts were intended to signal the car's autonomous action.

2.4 Limitations

One limitation of our study lies in the use of the simulator. Although it provides a reliable method for manipulating the experiment and a safe environment for testing still-experimental autonomous driving technology, there is a fair amount of difference in fidelity between the simulation and real-life driving. Conducting driving research in real traffic and road conditions would increase the reliability and validity of the data with regard to both attitudes and performance.

Second, the sample group is limited demographically: Participants are university students less than 30 years old. A wider range of participants—including newer drivers as well as elderly drivers with longer experience but possibly slower perception and reaction times—could yield an opportunity to more broadly generalize our findings or to produce different findings.

2.5 Procedure

Participants came to the driving simulator lab and, prior to driving, were given a brief description of the driving environment and signed an approved human subject consent form. After being acclimated to the simulator by driving 5 min on a practice course, participants drove for approximately 30 min on a 12-km test course. Participants were advised to drive safely and to obey traffic regulations such as speed limits, traffic lights, stop signs, etc. After completing the driving course, participants filled out an online questionnaire that assessed their overall driving experience and their reactions to dealing with the warning system. All participants were debriefed at the end of the experimental session.

2.6 Dependent variables

2.6.1 Attitudinal measures

Attitudinal measures were based on self-reported data on adjective items in the post-drive online questionnaire, and the order of the list of adjectives was randomized every time a new survey was run. The questionnaires were adapted from a published model from the CHIMe Lab at Stanford University that is used to measure driver attitude [10,11], with participants ranking each item on a ten-point Likert scale ranging from "Describes Very Poorly (=1)" to "Describes Very Well (=10)." With the self-reported rating scale, two indices, "emotional valence" and "machine acceptance," were created by averaging participants' ranking of adjectives. Both indices were tested for reliability using Cronbach's α, which implies that each set of items is closely related as a group.

The emotional valence index reflects responses to the question, "How well do the following words describe how you felt while driving?" The index was generated by averaging responses to the adjectives "anxious," "annoyed," and "frustrated" (Cronbach's $\alpha = 0.77$). This index was reversely coded so that higher scores were associated with a more positive emotional response.

The machine acceptance index reflects responses to the question, "How well do the following adjectives describe the car?" The index was generated by averaging responses to four adjectives: "intelligent," "helpful," "dominant," and "reliable" ($\alpha = 0.73$).

2.6.2 Behavioral measures

Safe driving behavior was objectively assessed by analyzing data collected from the driving simulator, including the following six items: collisions, speeding incidents, traffic light violations, stop signs missed, road edge excursions, and driving time.

3 Results

3.1 Driver attitudinal response

Emotional valence. Figure 3 depicts the analysis of the emotional valence index. A significant interaction effect emerged between two independent variables, the *why* and *how* messages, $F(1, 60) = 5.90$, $p < 0.001$. In the con-

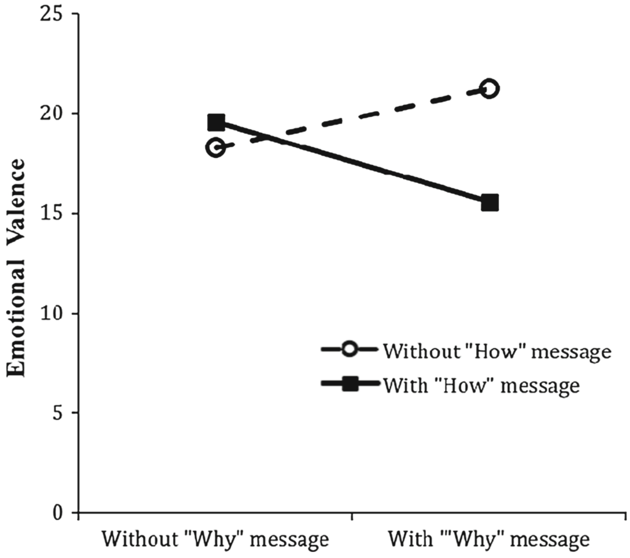

Fig. 3 Drivers' emotional valence. Y-axis corresponds to the mean value of the index, and a higher score indicates a more positive attitudinal response

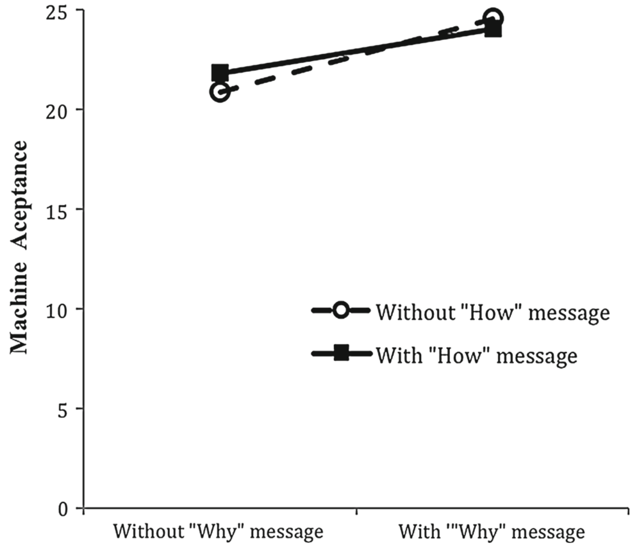

Fig. 4 Machine acceptance. Y-axis corresponds to the mean value of “machine acceptance,” and a higher score indicates greater trust in the automated system

text of auto-braking actions, people felt least positive about the *how* message when it was accompanied with the *why* ($M = 15.57,\ SD = 7.16$). They showed the most positive valence emotions when the *how* message was excluded ($M = 11.75,\ SD = 4.57$).

Machine acceptance. Figure 4 illustrates the analysis of driver acceptance of the automated system. There was a significant main effect from the independent variable, the *why* message, $F(1, 60) = 4.79$, $p < 0.05$. Drivers expressed greater system acceptance with messages that provided information about the driving environment ($M = 24.32,\ SD = 4.35$) than with messages that did not provide information about the driving environment ($M = 21.35,\ SD = 5.95$).

3.2 Driving behavior

Among various safe-driving performance measures, the only significant effect was upon the road edge excursions.

Figure 5 shows the correlation between the type of alert and road edge excursions. There was a significant interaction effect between the type of message and this measure of driving performance, $F(1, 60) = 6.92$, $p < 0.01$. When participants were not told the *how* message, the *why* made no difference. When participants were told the *how* message without the *why*, they drove worse ($M = 2.81,\ SD = 1.68$). Their safest driving performance was when they had both *how* and *why* messages ($M = 1.06,\ SD = 0.92$).

4 Discussion

Results show that, in a semi-autonomous driving situation, the type of car-to-driver communication about the car’s

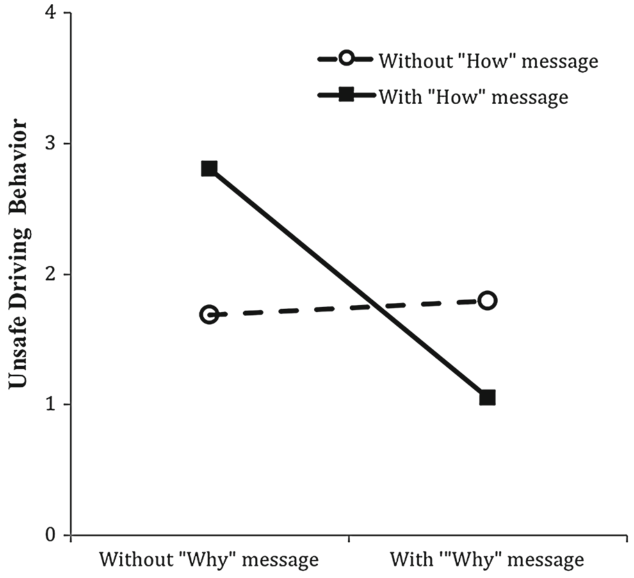

Fig. 5 Unsafe driving behavior. Y-axis indicates the number of road edge excursions

impending actions (*how* information and/or *why* information) has a significant effect upon drivers’ attitudes and behavior.

4.1 Information that conveys machine behavior and situational reasoning: the $how + why$ message

Initially, we hypothesized that people would prefer the *how+why* message and that it would improve driving behavior. Our rationale was based on by the well-known fact that users in the traditional realm of human–machine/computer interaction are typically more comfortable being informed of the system’s operating status when the information also conveys the reason for operating actions. In our study, surprisingly, car-to-driver communication that conveyed both the car’s actions and the reason for those automated actions affected driver attitude negatively. When people were told both how and why the car was about to act on their behalf, they felt anxious.

Now we understand that our hypothesis was based on insufficient consideration of the driving context. Imagine one scenario: a car gives you a vocal alert saying, “Car is slowing down because of obstacles ahead!” At this moment, the driver must process two types of information: the machine’s status and the situational status. Then the driver wonders, “What should I think about first?” As a consequence, it is difficult for the driver to determine what must be comprehended and responded to. With overloaded input, the driver could easily misunderstand signals. We believe that the *how+why* message, which creates the greatest cognitive load of the four experimental conditions, may be too much information to process while driving. The multiple simultaneous messages cause confusion and lead to an increased anxiety level.

Reeves and Nass, in their Media Equation theory [12], argue that too much or too little information may devalue the content of communication. The result of this problem of quantity is user frustration. One solution they proposed is to use properly abbreviated messages that both parties have prior agreement on. We call this "concurred abbreviation."

In signal detection theory, there is the concept of the receiver operating characteristic curve (ROC). The concept is that possibilities of mis-alarm (delivering an untimely signal) and false alarm (delivering the wrong signal) are increased when multiple alerts occur simultaneously [13]. In contrast to desktop computer scenarios, driving requires a quick reaction time to determine what is happening and respond to changes in the environmental situation. Thus, by making drivers rely on overloaded perception and cognition, complex signals can easily make drivers misread timely information, causing increased anxiety. This notion supports our finding that the combined *how+why* alert can cause negative effects.

Even though it was perceived negatively by drivers, the combined *how+why* message contributed to safer driving by minimizing off-lane excursions. By providing both the situational and operational contexts, the combined message helps drivers to maintain responsibility for controlling the vehicle when manual and autonomous controls co-exist. This fact underscores an important lesson for designers: the consumer appeal of the product does not always correlate with high performance and satisfaction in actual use. We should anticipate a potential design trade-off between attitudinal preferences and driving safety.

4.2 Information that conveys only machine behavior (automation-centered communication): the *how* message

The message reporting only the machine behavior caused drivers to perform the worst: drivers tended to drift out of their lane. This suggests that the *how* message, manifested as indicating the explicit behavior of the car such as "Car is braking," reinforces the idea that the responsibility over the car is held by the automated system, not the driver. The driver then tends to take a passive role, adopting the notion that "I don't have to react because the car will act on my behalf."

This finding can be explained by the concept of "locus of control": whether drivers feel that they (an internal determinant) or the automated system (an external determinant) are mainly responsible for the behavior of the vehicle [6]. Even though the car was only responsible for braking, the *how* message that created an external locus of control might have led a driver to assume a passive position relative to the automated system. As a result, this passive role caused the driver to fail to maintain a sense of control even over steering and lane-keeping, which led to decreased safe driving performance. It is notable that the failure to maintain the sense of longitudinal control had an impact on maintaining lateral control. In semi-autonomous driving situations, the driver and car should be seamlessly cooperative in the task of control transfer.

Another reason that drivers performed worse when given only the *how* message may have to do with situational awareness. The natural sequence of a driver's process is to perceive first, then react. In our study frame, the *how* explanation causes drivers to remain in a cognitive state of uncertainty: "Okay, I understand that this car is going to slow down for safety purposes, but why exactly?"

Endsley's model [14,15] is often cited to explain the concept of situation awareness (SA). In his definition, SA is the ability to perceive the related elements of the environment, to comprehend the given situation, and to anticipate the future status. This notion of SA is a crucial component in safe driving. We can surmise that the lack of situational reasoning in the *how* message explains why people drove unsafely under the *how*-only message.

Why would the *how* message, delivered without why information, have a negative effect on situation awareness in driving? Although the *how*-only alert was not as disliked as much as the combined *how+ why* message, drivers still didn't like the *how*-only alert. Perhaps this reaction stems from the idea that it is inconsiderate to provide messages that don't help the user know what to do.

In seeking better communication models between humans and machines, Reeves and Nass accentuate that designing polite machines is important because we humans are polite to machines, and if the machine's behavior fails to be polite in return, the failure is considered offensive [12]. Explaining the car's behavior, "Car is braking," without explaining the reason may be "impolite" even though it is accurate. The driver might get nervous not because the information was inaccurate but because the message was not helpful.

An alternative explanation for why the *how*-only message was disliked more than the *why*-only message is that the *how* information is redundant because it describes actions that the car is about to take, and the very action itself is a message; drivers might perceive the message + action as belaboring the point unnecessarily.

4.3 Message that conveys only the situational reasoning for automation (context-centered communication): the *why* message

Our results showed that drivers preferred receiving only the *why* information, which created the least anxiety (Fig. 3) and highest trust (Fig. 4). In contrast, the combined *how+why* information was the worst from a driver perspective. The succinct *why* message, such as "Obstacle ahead!" includes information about the environment that the car is about to encounter and offers the reasoning for the car's imminent actions.

This finding supports Reeves' and Nass's idea proposing to use properly abbreviated messages agreed upon by two parties [12]. As long as the succinct *why* message is well understood beforehand to encompass the occurrence of the car's autonomous action (concurred abbreviation), we designers may be able to enhance the car–driver relationship.

The amount of time that it takes to hear and cognitively process the *why*-only message is roughly the same as it takes the process the *how*-only message. The fact that performance and attitude are better in this condition suggests that the *why* message is more salient and that people process it quicker. For optimal driver response, there should be a balance between processing time and the information being processed. Information overload affects data quality. On the road, drivers continuously cycle through the information process of perceive–comprehend–anticipate. Drivers may find the data valuable and credible when the data content satisfies their cognitive need. For this reason, the *why* alert is more important to drivers and safer because drivers can anticipate upcoming events; the back-channel *why* cue helps drivers coordinate their reaction to the situation. As a consequence, the *why* message provides a useful way to enhance the interaction between the driver and the in-vehicle information system. These findings are consistent with previous research by Young and Stanton [16].

5 Conclusion

This research suggests an interaction model for cars to communicate with the human driver in the context of semi-autonomous driving. When autonomous driving coexists with manual driving, safety still demands high attention from the human operator. The main design lesson is the importance of providing the appropriate amount and kind of information to the driver. Too much information overwhelms the driver, even when the information is helpful to performance. The wrong kind of information can add to cognitive load or decrease the driver's sense of responsibility for the driving performance. When the possibility of transfer of control exists between the human driver and the car, the *how*-only message confuses the human operator about who is responsible, thereby causing unsafe driving behavior.

In summary:

1. *Why* information maintains good driving performance and is preferred by drivers.
2. *How* information without *why* information leads to dangerous driving performance.
3. *How*+*why* information can bother drivers but leads to the safest driving performance.

We conclude that both *how* and *why* information are needed for critical safety situations. Although including the reason for the car's behavior can incur a negative emotional response, it improves safety performance. Alternatively, the *why* message provides a moderate amount of information without causing a negative emotional reaction. The *why*-only message might be optimal when the car knows it is in a non-safety-critical situation.

Acknowledgments We thank Nissan Motor Cooperation, Japan, for supporting this research.

References

1. Bilger, B.: Has the self-driving car at last arrived? The New Yorker (2013). http://www.newyorker.com/reporting/2013/11/25/131125fa_fact_bilger?currentPage=all. Accessed 5 Feb 2014
2. DMV News Release (2012). http://www.dmvnv.com/news/12005-autonomous-vehicle-licensed.htm. Accessed 4 Feb 2014
3. DMV News Release (2014). http://www.dmv.com/blog/self-driving-car-regulations. Accessed 4 Feb 2014
4. Norman, D.: The Design of Future Things. Basic Books, New York (2007)
5. Norman, D.: The 'problem' with automation: inappropriate feedback and interaction, not 'over- automation'. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. **327**(1241), 585–593 (1990)
6. Stanton, N.A., Young, M.S.: Vehicle automation and driving performance. Ergonomics **41**(7), 1014–1028 (1998)
7. Walker, G.H., Stanton, N.A., Young, M.S.: The ironies of vehicle feedback in car design. Ergonomics **49**(2), 161–179 (2006)
8. Wiese, E.E., Lee, J.D.: Attention grounding: a new approach to in-vehicle information system implementation. Theor. Issues Ergon. Sci. **8**(3), 255–276 (2007)
9. Lee, J.D., Seppelt, B.D.: Human Factors in Automation Design, Handbook of Automation, pp. 417–436 (2009)
10. Takayama, L., Nass, C.: Assessing the effectiveness of interactive media in improving drowsy driving safety. Hum. Factors **50**(5), 772–781 (2008)
11. Takayama, L., Nass, C.: Driver safety and information from afar: an experimental driver simulator study of wireless vs. in-car information services. Int. J. Hum. Comput. Stud. **66**(3), 173–184 (2008)
12. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and News Media like Real People and Places. Cambridge University Press, Cambridge (1996)
13. Parasuraman, R., Anthony, J.M., Peter, A.H.: Fuzzy signal detection theory: basic postulates and formulas for analyzing human and machine performance. Hum. Factors **42**(4), 636–659 (2000)
14. Endsley, M.R.: Measurement of situation awareness in dynamic systems. Hum. Factors **37**, 65–84 (1995)
15. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors **37**, 32–64 (1995)
16. Young, M.S., Stanton, N.A.: Mental workload: theory, measurement and application. In: Karwowski, W. (ed.) International Encyclopedia of Ergonomics and Human Factors, vol. 1, pp. 507–509. Taylor and Francis, London (2001)